

**TRANSCRIPTOME PROFILING OF TEMPERATE AND
TROPICAL ADAPTED CARROT (*Daucus carota* L. subsp
sativus) CULTIVARS TO DECODE THE CANDIDATE
GENES RELATED TO ROOT ADAPTABILITY AND
FLOWERING**

CHAITRA C. KULKARNI



**DEPARTMENT OF BIOTECHNOLOGY AND
CROP IMPROVEMENT
COLLEGE OF HORTICULTURE, BAGALKOT- 587 104
UNIVERSITY OF HORTICULTURAL SCIENCES
BAGALKOT – 587 104**

FEBRUARY, 2022

**TRANSCRIPTOME PROFILING OF TEMPERATE AND
TROPICAL ADAPTED CARROT (*Daucus carota* L. subsp
sativus) CULTIVARS TO DECODE THE CANDIDATE
GENES RELATED TO ROOT ADAPTABILITY AND
FLOWERING**

*Thesis submitted to the
University of Horticultural Sciences, Bagalkot
in fulfilment of the requirements for the
Degree of*

DOCTOR OF PHILOSOPHY

In

GENETICS AND PLANT BREEDING

By
CHAITRA C. KULKARNI
UHS17PGD214

**DEPARTMENT OF BIOTECHNOLOGY AND
CROP IMPROVEMENT
COLLEGE OF HORTICULTURE, BAGALKOT- 587 104
UNIVERSITY OF HORTICULTURAL SCIENCES
BAGALKOT – 587 104**

FEBRUARY, 2022

**DEPARTMENT OF BIOTECHNOLOGY AND
CROP IMPROVEMENT
COLLEGE OF HORTICULTURE, BAGALKOT
UNIVERSITY OF HORTICULTURAL SCIENCES, BAGALKOT-
587 104**

CERTIFICATE

This is to certify that the thesis entitled “**TRANSCRIPTOME PROFILING OF TEMPERATE AND TROPICAL ADAPTED CARROT (*Daucus carota* L. subsp *sativus*) CULTIVARS TO DECODE THE CANDIDATE GENES RELATED TO ROOT ADAPTABILITY AND FLOWERING**” submitted by **Ms. CHAITRA C. KULKARNI** bearing **ID. No. UHS17PGD214** for the degree of **Ph.D. in GENETICS AND PLANT BREEDING** to the University of Horticultural Sciences, Bagalkot is a record of research work done by her during the period of her study in this university under my guidance and the thesis has not previously formed the basis for the award of any degree, diploma, associate ship, fellowship or other similar titles.

Place: BAGALKOT

Date: FEBRUARY, 2022

**(SARVAMANGALA S. CHOLIN)
MAJOR ADVISOR**

Approved by

Major advisor:

SARVAMANGALA S. CHOLIN

Members: 1._____
(H. B. PATIL)

2._____
(R. K. MESTA)

3._____
(MALLIKARJUN AWATI)

4._____
(S. R. MULLA)

ACKNOWLEDGEMENT

The work presented in this thesis was possible because of unconditional support of many people, my gratitude to them.

*With sincerity and immense pleasure, I would like to express my deepest gratitude to my advisor, **DR. SARAMANGALA S CHOLIN** Assistant Professor of GPB, University of Horticultural Sciences, Bagalkot. For the continuous support during my Ph.D. research over the past four years. **I am most grateful for her patience, motivation, and immense knowledge** and participation in various scientific discussions over the course of my Ph.D. Without her continuous motivation and encouragement, this work wouldn't have become a reality. I am most grateful to her for her patience during tough times. I spent 6 months at her home for thesis preparation during pandemic lockdown period, since college and hostels were closed. The care and professional time spent with her during thesis manuscript preparation time has created a great impact on my way of thinking towards challenging research works. Most thankful for all the efforts, to train me in bioinformatics, data analysis and interpretation of complicated results. She gave me the freedom to think independently and the opportunity to grow as a scientist. Grateful for the thesis supervision, fruitful discussions and kind support at all stages of this thesis. She has gone above and beyond her responsibilities as an advisor during research works and thesis preparation, and I could not have asked for a better person than her to lead me on this journey. I could describe her as a person with calm mind and positive thoughts. As a student of her, I have blessed and gifted feel at heart. Throughout the Ph.D. program, she was always quick to respond to all of my questions as well as fixing any problems I encountered.*

*Many thanks to members of my advisory committee - **Dr. H. B. Patil Prof. and former dean COH Bagalkot, Dr. R. K. Mesta Special officer, PCMC UHS, Bagalkote, Dr. Mallikarjun Awati**, Assistant professor (Crop physiology), COH Bagalkot and **Dr. S.R. Mulla** Assistant professor (Plant Biotechnology), HREC, Vijayapur for their suggestions and support throughout my doctoral program.*

*I am very grateful for the support of **Dr. H. B. Patil** for his every support during critical decisions related to extending working time of laboratory and providing financial support from college to take workshops related to bioinformatic analysis. My special thanks from inner core of my heart to **Dr. R. K. Mesta** for his constant support*

to carry out research works at molecular biology lab of **Dept. of Plant pathology** even after college hours and providing lab facilities to carry out **qPCR works**. I emphatically thank to **Dr. Mallikarjun Awati** for his every support and meticulous suggestions during flowering experiment at cold room.

My heartfelt sincere gratitude to **Dr. Prabhuling G.**, Professor and Head, Department of Biotechnology and Crop Improvement, College of Horticulture, Bagalkot for his timely help and guidance, and also for giving me the opportunity to conduct another important part of my research at the **phytotron** facility provided by **KBITS, Department of IT, BT, Govt. of Karnataka**. Which gave a different motto and passion for flowering experiment in controlled cold room chamber.

I will be forever thankful to **Dr. M. S. Kulkarni** former PG Dean and present Director of Education, UHS, Bagalkot for his valuable guidance and moral support which always helped me to focus on good work.

It is my prime responsibility to acknowledge **Dr. Indiresk K.M. VC, UHS Bagalkot** for his continued support and encouragement to carry the academic and research works in a smooth and cordial way.

My special acknowledgements to **Mr. Akhilesh Bajpai and Dr. Kshitish Acharya from Shodhaka Life Sciences**, supported madam and me throughout our bioinformatic analysis of root and FPL tissue transcriptome and helped us to acquire the knowledge related to computing resources needed for my work. Then, **Mr. Ajay Cholin** for extending his hand for **cloud computing** to handle the FPL tissue transcriptome.

My heartfelt gratitude to **Mr. Satish Patepur, Asst. professor, Dept. of Fruit science, MHREC, UHS Bagalkote** for providing the field and polyhouse facility during flowering phase experiment. **Mrs. Laxidevamma T.N. and Dr. Prabhu Ajjappalavar** for maintaining the field for stability analysis genotypes at Arabhavi and Devihosur locations. I would like to convey my special thanks for Mr. Prakash H.M who unconditionally provided polyhouse facility for my root tissue growth at vegetative phase.

I would like to extend my deep thanks and gratitude to all teachers, **Dr. Satish D.** Assistant Professor of GPB, UHS, Bagalkot. and **Mrs. Rekha Chittapur** Assistant Professor of GPB, UHS, Bagalkote for all their co-operation and help, necessary moral

support throughout degree program, **Dr. Ambika D. S.** Assistant Professor of Plant pathology, UHS, Bagalkot. **Dr. Manjunath Ramanna**, Assistant Professor of Microbiology, UHS, Bagalkot **Dr. Shivyogi Ryavalad** Assistant Professor of Seed Technology, UHS, Bagalkot and **Dr. Mahesh Y.S.** Assistant Professor of Plant pathology, UHS, Bagalkot. **Dr. Manjunath Hubballi** Assistant Professor of Plant pathology, UHS, Bagalkot for their constant supports at every accomplishment during academic and research works.

I would also like to express my sincere gratitude to **Dr. Balaji Kulkarni**, Dean, COH Bagalkot, for every good vibration he created during my doctoral degree days.

Words alone cannot express my deep sense of gratitude for the affectionate "At Home" memories of **Sri. Ningappa M. Shebannavar** and **Smt. Shrikanta N. Shebannavar** family members, will be cherished by me in the years to come. Their endless support, comfort, and encouragement have been the foundation to my success, and I could not have made it through this difficult process without them. The smile you brought on my face at the end of this stupendous thesis task is unforgettable in my entire life.

I gratefully acknowledge **Dr. Sunil N. Shebannavar**, for being there on my side and supporting with your expertise suggestions and ethical concerns. Let me always count my blessings from you. You have influenced my life in such a positive way, you are truly an inspiration and be remembered as a best human being. Thanks for being so generous thoughtful and true.

This acknowledgement of mine would undoubtedly be incomplete if I do not include these worth mentioning kiddos, **Mayur S. Shebannavar**, **Kritika S. Shebannavar** for all support, love, soft and sweet memories they have left behind in home and heart. You both inspire me to be my best, watching you both doing important transcriptome analysis under madam advise has been my greatest joy. I can never forget their collaboration and friendship and fun during thesis preparation time for cheering me up and the fun! I am truly fortunate to have been in the company of you both with brilliant minds and I am sure they are destined to achieve great success in their respective fields.

Words are not enough to express my gratitude towards my family. Without the support of my parents **Sri. Chidambar P. Kulkarni** and **Smt. Narmada C. Kulkarni** and

my dearest sister **Sushmita C. Kulkarni**. I wouldn't have reached this stage in my life without their blessings and hard efforts. Their support, sincerity, encouragement, patience, prayers and mellifluous affections that inspired me to achieve my goal and pursue my dreams. Grateful for their every support and sacrifices throughout my graduate career. It would be unjust to leave mentioning my uncle, **Sri. D.P. Vaidya** and aunt **Smt. Laxmi D. Vaidya** also sister **Anagha D. Vaidya** for supporting me spiritually and affectionately throughout my life. I would like to submit my sincere gratitude to **Sri. Ambareesh Karadi**, for his unconditional supports and attentions for the societal responsibilities.

I owe special thanks to **AgriGenome Labs Pvt Ltd, Kerala** who generated RNA sequencing data for both storage root and FPL tissue.

I would like to extend my deepest thanks to my beloved friends **Archana T. Sathyan** and **Krishnaja R. Nair** for their great help, valuable suggestions, considerable advices and support in all step of my research. I should thank them for sharing their thoughts and feelings with me and being interested in my opinion, helped me at different stages of my research work. Blessed for having you both as my support system and thanks a lot for always pushing me to go the extra mile.

I would also like to specially thank my funding sources; **Vision Group of Science and Technology, Govt of Karnataka, KSTePS, (DST-Ph.D. Fellowship Govt. of KNK)**. Also, **DBT (BioCARE)**, Govt of India. For funding to research work without which this work would not have been possible.

Its immense pleasure to extend huge, warm thanks to my former and present lab mates, **Santhosh, Chandan, Radhika Yadav, Soujanya, Divya, Suchita, Sahana** for valuable supports, and senior **Pavan Kumar Dumble** for his expertise suggestions during qPCR works and other juniors **Teju, Anita, Ananya**, who gave hands to me during all critical work times of research. Seniors, **Shruti and Chaitra A.P** for always being my side to keep my vibrations high.

I am also thankful to lab Assistant **Miss. Vidhya Choudhari**, and field assistant **Mr. Krishna, Mr. Jagadish Kumbar and Mr. Basavaraj Hitnalli** for all caring and supportive attitude during my research work along with, **Mr. Dilawar and many other labours of UHS Bagalkot**.

*I take this opportunity to express my gratitude to **Mrs. Girija B. Yandigeri, and Dr Girish Ganjihal, Asst. Librarians** and other faculty members for their support and help in Sarvajana library.*

*Last but not the least, am sincerely grateful to the whole **staff at UHS, Bagalkote** for offering me the opportunity to have a good quality education and to provide support to research work. From where, am starting my first step in my professional career.*

*It's my immense pleasure to acknowledge my two external evaluators. First, **Dr. S. Thirumeni, Professor of Plant Breeding and Genetics, Pandit Jawaharlal Nehru College of Agriculture and Research Institute, Karaikal** for all the inspirations, his words and blessings uplifted my interests in research which lasts forever. Second, **Dr. Sheeja T.E. Principal Scientist Division of Crop Improvement ICAR-Indian Institute of Spices Research (IISR)** for every suggestion in thesis write-up to increase the value of my thesis.*

*Lastly, I would again like to whole heartedly thank **Dr. Sarvamangala S. Cholin** madam. Your vast knowledge and experience of various research works in plants; for conducting my **M.Sc.** and **Ph.D.** work including population structure, RNA-seq bioinformatic analysis and phenotypic data analyses for candidate gene discovery and marker development and molecular breeding helped me to become a better researcher than I am. It has been such a pleasure working with you, and I thank you all for your commitment supervision of research and many inspiring discussions.*

Epilogue:

I can't talk butterfly language with caterpillar people.

I fly, I don't crawl!

Six and half years of my journey was fantastic. This has been a journey of learning, re-learning and unlearning tonnes of things. I would like to pen down something here.

Please neglect interpersonal elements during scientific discussions. Positive encouragement works better than criticism. Advises should be doors opening not doors closing. Because "Diversity is closely related to everything that brings development of humankind". It takes no efforts to destroy someone's hopes or hard work but one

encouragement lasts forever. Behind passion there are people with the courage to try. I don't believe, I have special talents. But I have persistence. I was taught that the way of progress was neither swift nor easy. It's a lifetime process to reach the level of excellence.

I am proud of my mentor and myself for how we were handling this research these past six years. Now, it's time to celebrate our strengths. Nothing will ever replace the human contact between learner and teacher. It takes lot of work and dedication to be a really, really good researcher. We were constantly challenged by many things during our works. Every challenge was really coming up well with good questions and answers. It's the best kept secret. Science never ceased to amaze us. In a number of our works during this thesis preparation, we are just narrating a tale for whoever is interested.

I remember the words said by Ratan Tata "I admire people who are very successful, and talented. But if that talent is with too much ruthlessness, I appreciate the talent, but I can't respect him".

I see myself as the one who never give up." I know that the most joy in my life has come to me from my research on carrot" The ultimate source of happiness is within me. For me, plant life and human health are mysteries. I want to have some deeper insight into both.

-Wish me good luck.

BAGALKOT
FEBRUARY, 2022

CHAITRA C KULKARNI



Sincerely Dedicated to
My Guru
Dr. Sarvamangla S. Cholin,
Parents and Sister

“० j vÆÃj zAv, UÄgÄ ०Ã½zAv”

CONTENTS

Chapter No.	Chapter particulars	Page No.
	CERTIFICATE	iii
	ACKNOWLEDGEMENT	iv
	LIST OF ABBREVIATIONS	xv
	LIST OF TABLES	xvi
	LIST OF FIGURES	xx
	LIST OF PLATES	xxiv
	LIST OF APPENDICES	xxv
1.	INTRODUCTION	1-6
2.	REVIEW OF LITERATURE	7-46
	2.1 Economic importance of carrot	7
	2.2 Taxonomy, Phylogeny and Distribution	8
	2.3 The domestication syndrome	10
	2.4 Secondary storage root development	11
	2.5 Two major cultivated carrot types	11
	2.6 Floral Development and Reproductive Biology	16
	2.7 Initiation of Flowering - Juvenility, Vernalization and Photoperiodic Requirements	17
	2.8 Flowering and Pollination	18
	2.9 Flower induction by the photoperiod pathway	21
	2.10 Regulatory network of flowering in long-day biennial carrot	21
	2.11 Carrot nuclear genome	24
	2.12 Gene Prediction and Annotation in WGS	26
	2.13 Carrot genome sequencing and assembly	26
	2.14 Comparative Analysis and Genome Evolution	28
	2.15 Transcriptomics studies in carrot (<i>Daucus carota</i> L.)	29
	2.16 Validation of RNA-Seq Results with Quantitative Real-Time PCR (qRT-PCR)	39
	2.17 Assessment of stability, adaptability and yield performance by Eberhard-Russel test of stability analysis of several genotypes	42

Chapter No.	Chapter particulars	Page No.
3.	MATERIAL AND METHODS	47-86
	3.1 Selection of plant material for RNA-seq	48
	3.2 Experimental design	50
	3.3 Workflow of Hiseq Illumina and generation of RNA-seq raw reads	52
	3.4 Computational analysis for Transcriptome assembly and Transcript quantification	56
	3.5 Identification of DEGs	57
	3.6 Functional annotation of DEGs across cultivars and Gene ontology of DEGs	66
	3.7 Gene regulatory Network (GRN) analysis/Co-expression analysis	72
	3.8 Transcriptional factor identification for Root and FPL tissues from <i>de novo</i> assembled unigenes	73
	3.9 Genetic variant analysis root and FPL tissue transcriptome	74
	3.10 Identification of candidate genes for root adaptability and flowering related genes	75
	3.11 Stability analysis of seven superior temperate/tropical adapted genotypes across seasons and locations	83
4.	EXPERIMENTAL RESULTS	87-164
	4.1 Root adaptability/root plasticity parameters of temperate and tropical carrot cultivars	88
	4.2 Generation of raw reads by transcriptome profiling of roots and FPL tissues in temperate and tropical cultivars	90
	4.3 Cleaned raw reads root and FPL transcriptome used for assembly	91
	4.4 Assembly/alignment of root (Reference based and <i>de novo</i> -based) and FPL tissue (only <i>de novo</i> -based) transcriptome	92
	4.5 Identification of DEGs for root transcriptome across cultivars in reference-based assembly	96
	4.6 Functional annotation of DEGs and gene functional classification	98
	4.7 Gene ontology (GO) categorization	100

Chapter No.	Chapter particulars	Page No.
	4.8 Gene set enrichment analysis (GSEA)	102
	4.9 <i>De novo</i> assembly	105
	4.10 Identification of differentially expressed genes by <i>de novo</i> assembly	112
	4.11 Functional annotation and GO classification, GO term enrichment analysis (Fisher's exact test), KEGG enrichment for DEGs across temperate and tropical cultivars root transcriptome	115
	4.12 Transcriptional Factors (TF) identified in root transcriptome in the present study	126
	4.13 Genetic variant analysis	130
	4.14 Gene Regulatory Network (GRN) analysis for storage root development	135
	4.15 Identification of Candidate genes for root adaptability and flowering related traits	136
	4.16 Validation of DEGs identified storage root by RNA-Seq results with quantitative real-time PCR (qRT-PCR)	142
	4.17 Expression pattern of flowering related DEGs between temperate and tropical cultivars by qPCR	147
	4.18 Gene Regulatory Network (GRN) analysis for flower transition between temperate and tropical carrots	148
	4.19 Assessment of stability, adaptability and yield performance of carrot (<i>Daucus carota</i> sub spp. <i>sativus</i>) cultivars in North Karnataka (tropical region)	148
5.	DISCUSSION	165-210
	5.1 Phenotypic plasticity of temperate and tropical and carrots	166
	5.2 Transcriptome profiling of root and FPL tissues	167
	5.3 Unigenes identification from <i>de novo</i> assembly in root and FPL transcriptome	171
	5.4 Identification of Differentially expressed genes (DEGs)	172
	5.5 Functional Annotation	174
	5.6 GO classification and gene set enrichment analysis (GSEA)-Root tissues	176
	5.7 Transcriptional Factors (TF) identified in root transcriptome	185

Chapter No.	Chapter particulars	Page No.
	5.8 Genetic variant analysis	189
	5.9 Gene regulatory network (GRN) analysis for storage root development	191
	5.10 Candidate genes related to root adaptability and flowering identified in the present study	195
	5.11 Validation of RNA-Seq Results with Quantitative Real-Time PCR (qRT-PCR)	204
	5.12 Assessment of stability, adaptability and yield performance of carrot (<i>Daucus carota sub spp. sativus</i>) cultivars	206
6.	SUMMARY AND CONCLUSIONS	211-216
7.	REFERENCES	217-262
	APPENDICES	263-277

LIST OF ABBREVIATIONS

RNA	: Ribonucleic acid
cDNA	: Complementary DNA
RNA-seq	: RNA Sequencing
Log2FC	: Logarithmic fold change
FDR	: False discovery rate
FPL	: Floral primordia and leaf
A	: Adenine nucleotide
G	: Guanine nucleotide
C	: Cytosine nucleotide
T	: Thymine nucleotide
U	: Uracil nucleotide
P value	: Probability value
DEGs	: Differentially expressed genes
GO	: Gene ontology
BP	: Biological Processes
CC	: Cellular Components
MF	: Molecular functions
TF	: Transcriptional factors
SNP	: Single Nucleotide polymorphism
EST	: Expressed Sequence Tags
SSR	: Simple Sequence Repeat
DNA	: Deoxyribose nucleic acid
SEm	: Standard error mean
%	: Per centage
qPCR	: quantitative Polymerase Chain Reaction
mRNA, tRNA, rRNA	: messenger RNA, transfer RNA, ribosomal RNA
miRNA	: Micro RNA
BLAST	: Basic Local Alignment Tool

LIST OF TABLES

Table No.	Title	Page No.
1.	Statistics of the “DC-27” and DH1 carrot genomes and gene predictions	27
2.	Summary of <i>Daucus species</i> short read archives (SRA) data available in NCBI	30
3.	List of genes identified in carrot for various traits of interest globally as on 2021	45
4.	Descriptions of two cultivated types utilized for Transcriptome profiling by RNA-seq	49
5.	Details of experimental plant material and Hiseq Illumina platforms used for Transcriptome profiling (RNA-seq) at vegetative and reproductive phases	53
6a.	RIN value for root tissue of two cultivars in three replications subjected to RNA-seq	54
6b.	RNA integrity number (RIN) value for the leaf and floral primordia/Bud samples of two cultivars in triplicates	55
7a.	<i>Raw reads obtained from RNA-seq by paired-end (PE) sequencing (2x100bp) for root tissues of temperate and tropical cultivars in triplicates</i>	58
7b.	<i>Raw reads obtained from RNA-seq by paired-end (PE) sequencing (2x151bp) for FPL tissues of temperate and tropical cultivars in triplicates</i>	59
8a.	List of genes used for RNA-seq validation of root tissue transcriptome by quantitative real time PCR	78
8b.	List of genes used for RNA-seq validation of FPL tissue transcriptome by quantitative real time PCR	79
8c.	List of genes used for validation of flowering related genes by quantitative real time PCR	80
9.	List of reference genes used for qPCR experiment for validation of root and floral primordial RNA-seq carrot	82
10.	Stability parameters for Eberhart and Russell, (1966) model	84
11a.	Summary of climatic factors for multi-location trial conducted during the present investigation	85
11b.	List of elite genotypes used for multi-location/multi-season evaluation	86

11c.	List of traits for stability analysis of carrot cultivars used for multi-location/multi-season evaluation	86
12.	Morphological traits distinguishing tropical and temperate carrot cultivars	89
13.	Summary of input reads, filtered and trimmed reads of six samples of two cultivars of carrot root and FPL tissue transcriptome data	93
14	Summary of the cleaned reads aligned to the reference genome by TopHat v2.0.6	95
15.	List of DEGs identified and annotated for reference-based root transcriptome	97
16.	Gene functional classification of Up and Down regulated genes with corresponding enrichment score and gene counts extracted from DAVID v 6.8	99
17.	GO term enriched in root transcriptome for BP, CC, MF and KEGG pathway along with NES and gene counts for root transcriptome	104
18.	Details of Denovo Assembly and unigenes identified after CD-HIT_EST in Root and FPL <i>de novo</i> assembly	107
19.	Results of the carrot root and FPL tissue transcriptome assembly assessment using BUSCO	109
20a.	Alignment rate of individual library of root transcriptome with Trinity assembly	110
20b.	Alignment rate of individual FPL tissue transcriptome library with Trinity assembly	110
21a.	KEGG pathways identified across temperate and tropical cultivars root transcriptome	120
21b.	KEGG pathways identified across temperate and tropical cultivars FPL transcriptome	124
22.	Number and classes of transcriptional factors identified by unigenes and 2239 common DEGs in root tissues	127
23.	List of differentially expressed TF families identified among temperate and tropical FPL tissue DEGs	129
24.	The summary of perfect microsatellite types identified in root tissue transcriptome	131

25.	List of SNPs and Indels identified in root transcriptome across temperate and tropical	133
26.	The summary of perfect microsatellite types identified in FPL tissue transcriptome	134
27a.	Differentially expressed candidate genes identified for 'root development' in tropical carrot cultivar	137
27b.	Differentially expressed candidate genes identified for 'root development' in temperate carrot cultivar	137
28a.	Differentially expressed candidate genes identified for carbohydrate and lipid metabolism in tropical carrot cultivar	139
28b.	Differentially expressed candidate genes identified for carbohydrate and lipid metabolism in temperate carrot cultivar	139
29a.	Differentially expressed candidate genes identified for photosynthesis in tropical carrot cultivar	140
29b.	Differentially expressed candidate genes identified for photosynthesis in temperate carrot cultivar	140
30a.	Differentially expressed candidate genes identified for biotic stress responses in tropical carrot cultivar	141
30b.	Differentially expressed candidate genes identified for biotic stress responses in temperate carrot cultivar	141
31a.	Differentially expressed candidate genes identified for abiotic stress responses in tropical carrot cultivar	143
31b.	Differentially expressed candidate genes identified for abiotic stress responses in temperate carrot cultivar	144
32a.	Candidate genes involved in flowering and reproduction identified from tropical carrot FPL transcriptome	145
32b.	Candidate genes involved in flowering and reproduction identified from temperate carrot FPL transcriptome	145
33a.	Mean Squares due to different source of variation for various quality traits in carrot (<i>Daucus carota L</i>)-ANOVA for Eberhart and Russel Model-Across seasons	150
33b.	Mean value, regression coefficient (bi) and variation due to deviation (s^2_{di}) for seven carrot genotypes across seasons	153
33c.	Character wise classification of genotype for stability and adaptability under different environments across seasons	154

34a.	Mean Squares due to different source of variation for various quality traits in carrot (<i>Daucus carota L</i>)-ANOVA for Eberhart and Russel Model across locations	156
34b.	Mean value, regression coefficient (bi) and variation due to deviation (s^2_{di}) for seven carrot genotypes across locations	161
34c.	Character wise classification of genotype for stability and adaptability under different environment	162
35a.	Environmental index and phenotypic index for root productivity traits across seasons	163
35b.	Environmental index and phenotypic index for root productivity traits across locations	164

LIST OF FIGURES

Figure No.	Title	Between Page
Flow chart		
1	RNA-seq experimental work flow and computational analysis	48-49
Figures		
1.	Tapestation profile Qualitative analysis of RNA of six root samples from Temperate and Tropical cultivars confirmed before RNA-seq by bioanalyzer	52-53
2	Tapestation profile of RNA of leaf and flower bud of Temperate and Tropical cultivars in triplicates confirmed before RNA-seq by bioanalyzer	56-57
3	<i>De novo</i> transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity (Haas <i>et al.</i> , 2013)	64-65
4	Selection of most stable endogenous gene control for qPCR experiment	82-83
5	Number of up and down regulated DEGs identified in pair-wise differential analysis in Cuff-diff package of Cufflinks suite v1.3.0	98-99
6a	Visualization of total DEGs identified by reference-based root transcriptome between tropical and temperate carrot cultivars by volcano plot	98-99
6b	Chromosome-wise distribution of DEGs identified for reference-based transcriptome between temperate and tropical root tissues based on FDR (<0.50) and Log2FC (<-1 to >1).	98-99
7a	Gene Ontology terms identified for Biological Processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93).	100-101
7b	Enrichment tree showing relationship of Gene Ontology terms identified for Biological Processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root.	100-101
8a	Gene Ontology terms identified for cellular component (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID).	102-103
8b	Enrichment tree showing relationship of Gene Ontology terms identified for Cellular components (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root.	102-103

9a	Gene Ontology terms identified for ‘Molecular Functions (MF)’ associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID).	102-103
9b	Enrichment tree showing relationship of Gene Ontology terms identified for Molecular Functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root.	102-103
10a	Gene set enrichment analysis (GSEA) pathway terms identified for biological process (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID).	102-103
10b	Pathway tree showing relationship of GSEA terms identified for Biological processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root.	102-103
11a	Gene set enrichment analysis (GSEA) pathway terms identified for cellular component (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart.	104-105
11b	Pathway tree showing relationship of GSEA terms identified for cellular components (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root.	104-105
12a	Gene set enrichment analysis (GSEA) pathway terms identified for molecular functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID).	104-105
12b	Pathway tree showing relationship of GSEA terms identified for Molecular functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root.	104-105
13a	Gene Ontology terms identified for KEGG enrichment analysis associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID).	104-105
13b	Enrichment tree showing relationship of pathways identified for KEGG enrichment analysis associated with Tropical carrot root tissue in comparison to temperate carrot root.	104-105
14a	Comparison of contig lengths (bp) between trinity assembly and unigenes in <i>de novo</i> assembled root transcriptome of tropical and temperate cultivars	106-107
14b	Comparison of contig lengths (bp) between trinity assembly and unigenes in <i>de novo</i> assembled FPL transcriptome of tropical and temperate cultivars	106-107

15a	Summary statistics of DEGs identified by DESeq, edgeR tools and number of common DEGs selected in root tissue across temperate and tropical cultivars	114-115
15b	Visualization of total DEGs identified by <i>de novo</i> -based root transcriptome between Asiatic and Western carrot cultivars by volcano plot	114-115
16a	Summary statistics of DEGs identified by DESeq, edgeR tools and number of common DEGs selected in FPL tissue across temperate and tropical cultivars	114-115
16b	Visualization of total DEGs identified by <i>de novo</i> -based FPL transcriptome between temperate and tropical carrot cultivars by volcano plot	114-115
17a	Data distribution of un-blasted, blasted, mapped and annotated sequences over the whole data-set of 2239 common DEGs	116-117
17b	Data distribution of un-blasted, blasted, mapped and annotated sequences over the whole data-set of 2239 common DEGs	116-117
17c	Sequence similarity distribution of the BLAST hits for each common DEGs sequence in the NR database with a cut off E-value of 1.0 e-5 for all matching hsp	116-117
17d	Species distribution of the BLAST hits for each Blasted DEGs in the NR database	116-117
18	Distribution of Gene ontology terms in genes exclusively present in the assembly of tropical and temperate carrots root transcriptome (A) Biological process, (B) Cellular component, (C) Molecular function	118-119
19a	GO categorization of DEGs identified across tropical and temperate cultivars for biological process (BP) in root transcriptome based on <i>De novo</i> assembly	118-119
19b	GO categorization of DEGs identified across tropical and temperate cultivars for cellular component (CC) in root transcriptome based on <i>De novo</i> assembly	118-119
19c	GO categorization of DEGs identified across tropical and temperate cultivars for molecular function (MF) in root transcriptome based on <i>De novo</i> assembly	118-119
20a	Fisher's exact test for upregulated DEGs as test set and total DEGs as reference set for GO groups in root transcriptom	118-119
20b	Fisher's exact test for down regulated DEGs as test set and total DEGs as reference set for GO groups in root transcriptom	118-119
21a	GO categorization of DEGs identified across tropical and temperate cultivars for biological process (BP) in FPL transcriptome based on <i>De novo</i> assembly	126-127
21b	GO categorization of DEGs identified across tropical and temperate cultivars for cellular component (CC) in FPL transcriptome based on <i>De novo</i> assembly	126-127

21c	GO categorization of DEGs identified across tropical and temperate cultivars for molecular function (MF) in FPL transcriptome based on <i>De novo</i> assembly	126-127
22a	Total transcriptional factors identified for the unigenes (44856), up and down regulated DEGs among root transcriptome	126-127
22b	Transcriptional factors identified for root transcriptome from unigenes lists	126-127
23	Transcriptional factors identified for FPL tissue transcriptome from unigenes lists	128-129
24a	Percentage of different perfect SSRs distribution in root transcriptome	130-131
24b	Microsatellite distribution analysis and top abundant perfect SSRs found in root transcriptome	130-131
25	Identification of SNPs (Transition, Transversion) and Indels (insertions and deletions) in root transcriptome data	132-133
26	Microsatellite distribution analysis and top abundant perfect SSRs found in FPL transcriptome	132-133
27a	Identification of SNPs and Indels in FPL transcriptome	136-137
27b	Types of SNPs identified in FPL transcriptome	136-137
28a	Gene regulatory network showing the network of genes filtered for secondary storage root development in temperate carrot based on orthology search with <i>A. thaliana</i> and network drawn in STRING v. 11	136-137
28b	Gene regulatory network showing the network of genes filtered for secondary storage root development genes in tropical carrot based on orthology search with <i>A. thaliana</i> and network drawn in STRING v. 11.0	136-137
29a	RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in storage root transcriptome of temperate and tropical cultivars	146-147
29b	RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in storage root transcriptome of temperate and tropical cultivars	146-147
30	RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in FPL tissue transcriptome of temperate and tropical cultivars	148-149
31	qRT PCR validation of flowering related DEGs between temperate and tropical cultivars based on gene expression values	148-149
32	Heat map showing expression pattern of flowering related genes between temperate and tropical cultivars CCA1, TOC1 -8am samples, CCA2, TOC2 4pm samples, CCA3, TOC3- 8pm samples.	148-149
33	Identification of conserved network of flowering regulation genes critical for floral transition in carrot identified from AT orthology	148-149

LIST OF PLATES

Plate No.	Title	Between Page
1.	Vegetative phase experiment- Selection of matured storage roots for RNASeq	50-51
2.	Vegetative phase change and floral transition in carrot	52-53
3.	Reproductive phase evaluation at field	52-53
4.	Phytotron experiment for flowering induction in temperate and tropical cultivars	52-53
5.	Temperate cultivars floral primordia	52-53
6.	Tropical cultivars floral primordia	52-53
7.	Stability analysis of seven elite carrot cultivars at Bagalkot location (Northern dry zone of Karnataka)	86-87
8.	Stability analysis of seven elite carrot cultivars at Arabhavi location (Northern Transitional Zone)	86-87
9.	Stability analysis of seven elite carrot cultivars at Devihosur location (Northern Transitional Zone)	86-87

LIST OF APPENDICES

Appendix No.	Title	Page No.
I	Functional annotation of DEGs identified in reference-based root transcriptome (Functional gene sets)	263-277

1. INTRODUCTION

After discovering the concept of totipotency in phloem tissue of carrot (*Daucus carota* L.), it is considered as a model crop for tissue culture (Steward *et al.*, 1958). It has also become a model plant species for somatic embryogenesis and carotenoid accumulation studies. Carrot is the first plant species approved for molecular farming and biotherapeutics to treat various human genetic and infectious diseases. Carrot offers provitamin-A, antioxidants, minerals, and other vitamins with zero fat critical for optimal health. A medium-sized carrot (~60g) provides enough provitamin-A carotenoid to fulfill the adult Vitamin-A needs for a day (Simon, 1990). Hence, it is considered to be the low-cost and richest source of carotenoid among the plant and animal products to effectively fight against vitamin A deficiency disorder (VAD), the most damaging micronutrient deficiency in the world. Given its nutritional value and economic importance, carrot is considered the second most important vegetable crop next to potato (Spooner, 2019; Heywood, 2014). Over the past five decades, global carrot production has increased because of the growing awareness of the benefits offered by carrots as a rich source of vitamins and minerals, including provitamin-A.

Carrot (*Daucus carota* L.) belongs to one of the largest families of seed plants, *Apiaceae* (Judd *et al.*, 2016), with the diploid chromosome number of $2n=2x=18$ and an estimated genome size of 473 Mbp (Iorizzo *et al.*, 2016). Being a highly outcrossing crop species, it has a weak biological barrier to inter breeding, making *Daucus carota*, a highly variable species that is depicted at both genotypic and phenotypic levels. Cultivated carrot (*Daucus carota* subsp. *sativus*) is categorized among 25–40 related wild species in the genus *Daucus*.

Cultivated carrot types are classified into two groups: anthocyanin or eastern/Asiatic/tropical-type carrot (e.g., yellow or purple and domesticated orange carrots), and carotene or Western/temperate-type carrot (e.g., yellow, orange, or red color carrots). Western-type is a temperate adapted cultivar and is highly consumed domesticated cultivar type and characterized by a storage root with the highest carotenoid and sugar content. The taproot of Western/temperate carrots is typically cylindrical or tapered cylindrical and plants ensure fewer pubescent leaves, flowers

biennially (Que *et al.*, 2019). While Asiatic-type is a tropical adapted type, possess comparatively thicker, shorter, narrow, and conical roots and plants contain pubescent leaves, flowers annually, and are meager in carotenoid content (Que *et al.*, 2019). Selection pressure has improved beneficial traits like greater provitaminA carotenoid and higher sugar content in Western carrots than the Asiatic carrot (Ellison *et al.*, 2018). Genetic mechanisms that contribute to phenotypic changes associated with these two cultivated types of carrot remain unknown even though temperate/Western carrots are evolved from Asiatic/tropical carrots (Simon, 2008; Iorizzo *et al.*, 2016; Ellison *et al.*, 2018).

Carrot is a major root vegetable with a taproot system rich in carbohydrates, other nutrients, and water. Vascular cambium (VC) plays a major role during secondary storage development. It involves the establishment and maintenance of vascular stem cells such as secondary meristem, that divide in a periclinal direction (sometimes regarded as lateral or radial direction) and produces xylem tissue inwards and phloem outwards (Haung *et al.*, 2020a). In model plants and other root crops like radish, a gene regulatory network mediated by transcription factors controls dynamic changes in vascular cell development. These transcriptional networks are co-expressed with many stress-responsive genes (Haung *et al.*, 2020a; Huang *et al.*, 2020b). Molecular understanding of VC development patterning in carrot would help in genetic manipulation of the undesirable xylem core present in Asiatic cultivar in contrast to uniform and deeply pigmented xylem/phloem pattern of the Western cultivar.

The root is referred to as the “hidden half” of the plant; a strong and well-established root system plays an important role in the plant performance and yield in the case of crop plants (Lynch 2011; Paez-Garcia *et al.*, 2015). Thus, for developing plants suitable for growing in different agro-ecological conditions, it is necessary to study the molecular mechanisms underlying the expression of genes and other regulatory factors. Roots are essential for plant adaptation and productivity but are less studied compared to above-ground tissues due to the difficulty of observing them during the plant life cycle. The root system architecture is made up of structural features like root length, number, spread, and length of lateral roots which exhibit great plasticity. Hence, key

genes involved in root development are critical for proper plant stand in response to environmental changes (Waines *et al.*, 2007; Den Herder *et al.*, 2010).

Photosynthesis, transport, and accumulation of carbohydrates are the three essential physiological processes in any economically important crop. Still, how photoassimilates are transported from leaf sources to storage root sinks is unexpectedly diverse among different species (Huang *et al.*, 2020). Phloem loading in leaves and unloading at the sink are two restricted steps that play a pivotal role in this process and can significantly affect crop yield and plant productivity. A cultivar with a high ratio of "storage root" to "shoot" is known to have the highest carotenoid and sugar content in carrots (Iorizzo *et al.*, 2016; Machaj *et al.*, 2018).

An agricultural crop's, adaptability is characterized by its ability to respond positively to changes in environmental conditions. For instance, root plasticity is considered a key objective in the development of more productive roots in diverse environments. Phenotypic plasticity offers many ecological and physiological benefits (Schneider and Lynch, 2020). Carrot crop has been classified into temperate climate-adapted Western carrots and tropical climate-adapted Asiatic carrots that have an ancestral relationship with each other and with recent domestication history (Ellison *et al.*, 2018). Despite having wider adaptation to warmer climates, tropical carrots have been least explored in varietal improvement due to the lack of knowledge of the genetic architecture of tropically adapted carrots at the molecular level. Comparative transcriptome profiling of this wider climate adaptable tropical carrots with cool climate adaptable temperate type carrots provides a dynamic range of expression levels between these cultivars. It further helps for a better understanding of the adaptability and flowering-related candidate genes that are differentially expressed during the vegetative and reproductive phases of the carrot lifecycle.

Carrot is a biennial crop and its life cycle consists of two phases' viz., vegetative and reproductive phases. Since the seed is a propagating material and root is the edible part, both phases are considered to be important for farmers and breeders. Junttila, (1991) reported that the reproductive phase contains various ontogenetic stages. In the preliminary stages, the plant prepares for flowering with the shoot apical meristem (SAM) stimulation for stalk elongation. Hence, the shoot apical meristem (SAM) is

predominantly concerned with many organogenic processes accompanied by shoot morphogenesis (Wang *et al.*, 2019). While, the sub-apical meristem is the second most important site of cell division and elongation furnishing stem extension, eventually transforming into an inflorescence meristem, and beginning the complex process of flower development. Then, floral meristem identity genes regulate the formation of the flower. Once the flowering initiates, floral genes which govern the whorl formation in flowers are expressed, the structures of the whorls and their occurrence at the right position are regulated by homeotic genes (Wang *et al.*, 2019).

Current knowledge on carrot reproductive biology's molecular mechanism is inadequate compared to the other model plants despite powerful genetic research on carrots and their wild relatives over the last decades (Linke *et al.*, 2019). A possible reason for this is the prioritization of research on economically relevant traits such as yield, quality, or resistance breeding in the vegetative phase. A focus on reproductive biology is missing; however, it is important for breeders and seed producers. The current knowledge helps to introduce characters from late-flowering temperate cultivars to early flowering tropical germplasm in breeding programs with more accuracy. Nevertheless, in the future, the knowledge of the reproductive biology of carrots should be improved to benefit from the availability of comprehensive carrot genome data.

Flowering initiation is controlled by a series of physiological pathways acting in response to environmental cues and endogenous signals (Andres and Coupland, 2012). The transition from the vegetative phase to reproduction is a crucial developmental switch in the plant biological clock and the timing of this switch is vital for reproductive success. The decision of transition from juvenile to reproductive phase is genetically driven by the interaction of endogenous factors (e.g., age and developmental state) and environmental cues, light, and temperature (acting through photoperiod and vernalization pathways) are the most reliable seasonal cues (Amasino, 2010, Andres and Coupland, 2012). Plants anticipate upcoming changes in environmental conditions and regulate flowering time in plants (Shim and Jang 2020) with the help of the above-mentioned cues. Although bolting and flowering are important constraints to carrot production, only a few genetic or molecular studies have focused on this process. The

stage of growth when carrots seedlings are not responsive to low-temperature vernalization is known as juvenility. In carrots, the level of response to cold treatments is cultivar-dependent Atherton *et al.*, 1990; Hiller and Kelly 1979; Alessandro and Galmarini 2007; DiasTagliacozzo and Valio 1994).

The carrot genome severely lacks good-quality coverage of transcriptomics data from different tissues and developmental stages in vegetative and reproductive phases either in normal conditions or during abiotic and biotic stress treatments especially from the diverse genetic background cultivars such as temperate and tropical adapted types. To the best of our knowledge, the carrot genome does not contain comparative ‘omics’ data between two domesticated cultivar types *viz.*, Asiatic (tropical-type) and Western (temperate-type) carrots for economic tissues, either for vegetative or reproductive stages till today. Such data and platforms are extremely relevant to understanding domesticated orange root cultivars of Asiatic and Western cultivars and their functionality at the root and flower developmental stages to elucidate the key genes at two developmental phases. The identified various candidate genes in the present study can be utilized for the selection of economic traits.

Considering the above facts and the current need for research in carrots, the present investigation was planned to elucidate the biology, molecular and genetic determinants involved in secondary storage root development and adaptation in the matured edible taproot of carrots. The present research work was also aimed to interrogate the molecular mechanism involved in vegetative to flowering transition during the reproductive phase. Two contrasting, economically important carrot types, comprising of temperate region adapted-Kuruda cultivar and tropical region adapted carrot cultivar, UHSBC-23-1 (a promising selection of UHS Bagalkot), were used for the experiment. To achieve the objectives of the research, two economic tissues *viz.*, matured storage root at vegetative phase and floral primordia (and corresponding leaf tissue) during reproductive phases were subjected to transcriptome profiling. Identified candidate genes will be further exploited from desirable traits of temperate carrot to develop the tropical varieties or *vice versa* by ‘Genomics Assisted breeding’.

The details of the objectives are as follows.

1. Transcriptome profiling of 'root' and 'Floral Primordia & Leaf (FPL)' of temperate and tropical adapted genotypes of carrot by RNA-seq
2. Differential gene expression analysis and identification of candidate genes for adaptability and flowering related genes
3. Functional validation of selected significantly differentially expressed genes by qPCR
4. Stability analysis of seven superior temperate/tropical adapted genotypes across seasons and locations

2. REVIEW OF LITERATURE

The chapter reviews the literature on carrot genetic diversity, domestication, the genetic makeup of carrots also recent trends and advances in the improvement of carrots. The present research was focused on the comparison of the temperate and tropical adapted carrot cultivars for their two economic tissues (root and floral primordia) that were grown in warmer climatic regions (tropical or sub-tropical). Hence, the chapter also highlights the breeding targets to know the strategic features in both edible storage root development as well as flower development (reproductive biology) in temperate and tropical carrots under the tropical region. The review similarly discusses the recent findings in the comprehensive genomic data obtained from carrot set (Iorizzo *et al.*, 2016) and transcriptomics findings in carrot from previous carrot research in relevance to root adaptability and flowering pattern. The Gene Regulatory Network (GRN) involved in the plasticity of flowering time and recently characterized components of root development is remarkably considered.

2.1 Economic importance of carrot

Carrot production has increased progressively in the last 50 years (FAO 2017), with a three times expansion of production area and two times intensification in yield about successively six times greater rise in total production. The aforesaid upturns resulted in the 2.7-fold average global increase in per capita carrot production in the last five decades.

The production and availability of carrots have expanded fairly in the lead of the averages compared to 22 primary vegetables, such that carrot currently attributes 5.5% of the proportionate vegetable accessibility all over the world. This favorable enhancement was remarkably observed in Asia. Both production trends and economic rise in the significance of the carrot crop were particularly high in Asia. At present, carrots are among the top 10 vegetables, based on worldwide production records of primary vegetables, after tomatoes, onions, cabbage, cucumbers, and eggplant (FAO 2017).

Carrot germplasm comprises a wide collection of genotypic and phenotypic diversity that contributes to its wider adaptability (Chaitra *et al.*, 2020). The broad range

of phenotypic variation present in carrot crops is concisely exploited by breeders (Simon *et al.*, 2008) and exploration of genotypic variation is merely in preliminary stages due to the availability of high-quality whole-genome sequence information (Iorizzo *et al.*, 2016).

In the beginning, varietal improvement of carrot was comprehensively supported in temperate regions of Europe and Asia, but cultivars adapted to tropical and sub-tropical climates have imparted appreciably to strengthen the worldwide carrot production in the past half-century. Orange carrots are mainly consumed for their provitamin A and carotenoids. With the enhancement of carrot production in warmer climates, it has become possible to reduce the vitamin A deficiency that was prevalent in warmer climates (Tanumihardjo, 2012). Even though carrots are classified as a cool-season crop, the adaptation of carrots to warmer climates raises a fruitful indication for continued expanded carrot production and improvements for developing climate-resilient varieties now and in the future.

2.2 Taxonomy, Phylogeny, and Distribution

Swedish botanist Carl Linnaeus was a pioneer who classified the carrot originally in his book *Species Plantarum* in 1799 (Linnaeus, 1799). The cultivated carrot (*Daucus carota* subsp. *sativus*) is the most important member of the family Apiaceae having essential nutritional properties and economic importance (Rubatzky *et al.*, 1999; Simon 2000).

Cultivated carrot (*Daucus carota* subsp. *sativus*) is categorized among 25–40 related wild species in the genus *Daucus* depending on the classification. It is a part of a widely distributed and taxonomically complex family *Apiaceae* (*Umbelliferae*) containing 466 genera and 3820 species that is one of the largest families of seed plants (Spooner *et al.*, 2019).

Carrot is regarded as the second most important vegetable in the world after potato (Heywood 2014). Despite its universal prominence, classification, taxonomy, diversity, and the systematics of *Daucus* are continuously updated at the species, genus, and sub-tribal levels in the taxonomic classifications. Banasiak *et al.* (2016) described the relevant nomenclature of the genus *Daucus* comprising ca.40 species. Among

which, the genus *Daucus carota* is the ancestor of cultivated carrots (*Daucus carota* L. subsp. *sativus* Hoffm.). *Daucus carota* L. is similarly comprised of more than 60 intraspecific taxa making *Daucus* the most challenging species group in the *Apiaceae* (Heywood 1968a, b; Small 1978; Thellung 1926).

Cultivated and wild carrots are taxonomically classified in the same species *D. Carota* and all the constituent taxa under this maintains $2n=2x=18$ chromosomes and have a weak biological barrier to interbreeding, making *Daucus carota*, a highly variable species (Nothnagel *et al.*, 2000). *D. Carota* can certainly undergo widespread hybridization with commercial varieties of carrot and wild species spontaneously in the field with apparently no barrier to crossing (Hauser and Bjorn, 2001; Hauser, 2002; Rong *et al.* 2010).

The taxonomic classification provided by *Linneaus* (1799) of carrots is presented below.

- Kingdom: *Plantae*
- Phylum: *Spermatophyta*
- Subphylum: *Angiospermae*
- Class: *Dicotyledonae*
- Order: *Apiales*
- Family: *Apiaceae*
- Genus: *Daucus*
- Species: *Daucus carota*

The phylogenetic alliances among the species of *Daucus* also its close relatives have been simplified conspicuously in the last few years by a series of molecular studies using DNA sequences of plastid genes like *rbcl* and *matK*, plastid introns *rp116*, *rp16*, *rpoc1*; nuclear ribosomal DNA, internal transcribed spaces (ITS) sequences; and plastid DNA restriction sites ((Arbizu *et al.*, 2014, 2016a, b; Banasiak *et al.*, 2016; Weitzel *et*

al., 2014). A high-quality reference genome sequence was developed by Iorizzo *et al.*, in 2016 which can be efficiently utilized for understanding the evolution and genomic regions under selection united with domestication.

2.3 The Domestication Syndrome

Carrot domestication is a co-evolutionary process that involves genetic modification of a wild species to develop new plant structure and physiology that distinguish cultivated and wild forms of carrot and evolved as edible storage root crops. Cultivated carrot has been highly modified when compared to its wild progenitors. However, both cultivated carrots are grown worldwide and wild carrots are also found globally on all continents except Antarctica (Grzebelus *et al.*, 2011).

Genes governing all the domestication traits are mostly unknown in carrots. The term “adaptation syndrome” was used by Harlan (Harlan *et al.*, 1973) to describe traits spontaneously selected through an adaptive harvesting regime of cereals later, the phrase “domestication syndrome” was first conceived by Hammer (Hammer, 1984) a decade later. The key genes influencing important traits are uncertain and the extent to which new mutations versus pre-existing variation in wild plants contributed to the evolution of Western/temperate-type carrot has to be studied in detail.

The domestication syndrome of carrot (*Daucus carota* subsp. *sativus*) is mainly directed towards edible storage roots concerning root color, shape, size, and flavor. The cultivated carrots are characterisedstorcharacterized with smooth and juicy storage roots with increased carotenoid, anthocyanin, sugar contents (Rong *et al.*, 2014) also relatively large, unbranched (loss of lateral branching), biennial growth habit, and variation of root shape. Among all these traits, only the *Vrn1* locus is a probable target to favor biennial growth habit all through the course of carrot domestication, it has been recently mapped on carrot chromosome 2 (Alessandro *et al.*, 2013). Yet, the gene has not been characterized. Continuously, after the aforementioned traits have been selected and fixed, the process of domestication has concentrated more consideration on a few other carrot improvement traits like better flavor, nutrition, uniformity, biotic and abiotic stress tolerance, and male sterility for hybrid cultivar development (Doebley *et al.*, 2006; Ellison, 2019).

By the pigmentation of the roots, carrot taproot comprised circular central core which contains the vascular tissues xylem and phloem. A pulpy outer cortex (phloem) and an inner core (xylem), phloem channels present in the root system transport sugars around the plant then the roots firmly move nutrients and water from the soil to the entire plant through xylem pathways. Pericycle cells essentially work to protect and maintain vascular tissue the xylem and phloem. Cortex is the outer portion of the carrot, which is made up of more phloem (Rubatsky *et al.*, 1999).

2.4 Secondary storage root development

Tissue-specific for food storage is usually made up of parenchyma cells. The ground tissue in carrots is specialized for sugar and starch storage. The sugars synthesized in leaves move down to the roots through vascular phloem cells to the parenchyma cells of ground tissue or cortex comprised of phloem tissue, conducting photosynthate from leaves to roots. Biennial carrots utilize the first summer for photosynthesis and root development. Then store the food in roots over the first winter in the course of vegetative root development. Then use these sugars in the reproductive phase to produce flowers and seeds during the second summer (Que *et al.*, 2019).

2.5 Two major cultivated carrot types

Cultivated carrots are broadly classified into two groups: anthocyanin or eastern/Asiatic/tropical-type carrot (ex. yellow or purple carrot and domesticated orange), and carotene or Western/temperate-type carrot (ex. yellow, orange or red carrot) in which taproot is typically cylindrical or tapered cylindrical and ensure fewer pubescent leaves (Que *et al.*, 2019). Selection pressure has improved beneficial traits like greater provitamin A carotenoid and higher sugar content in Western/temperate carrots than Asiatic/tropical-types (Ellison *et al.*, 2018). Though a completely xylem-free carrot is not found, many cultivars of Western/temperate types have small and deeply pigmented cores. High-quality Western/temperate-type carrot appears to have a lack of core because the color of the cortex and core are similar in intensity and is more desirable and fetch a higher price for the farmer. Asiatic/tropical-type carrots possess comparatively thicker, shorter, narrow, and conical roots. Asiatic/tropical-type carrot

plant usually contains pubescent leaves, flower annually, and are meager in provitamin-A carotenoid content (Que *et al.*, 2019).

Popularity for human consumption of Asiatic/tropical-type carrot at present is the least. Consumer preference has largely switched to the Western/temperate-type carrot (Heywood, 1983). It is believed that tropical type cultivated carrot originated in south Asia, particularly near Afghanistan only about 1100 years ago (Iorizzo *et al.*, 2013, 2016) regarded as the first domestication event of cultivated carrot. Historical records and experimental results support the indication of central Asia as the center of origin (Que *et al.*, 2019). However, there are many different viewpoints on the origin of Western/temperate-type cultivated plants. Until now, the origin of the Western/temperate-type cultivated carrot is not clear and has remained ambiguous.

The Western/temperate carrot origin from yellow Asiatic/tropical-type cultivars was demonstrated by Banga during 1957 and documented in Dutch paintings in support of the same. Western/temperate-type carrots were grown in the Netherlands during the 18th century and modern cultivated carrots are believed to be originated from old orange Dutch carrots (Banga, 1957a, b; Stein and Nothnagel, 1995; Simon *et al.*, 2008).

Heywood dis-agreed Western/temperate-type carrots evolution from Asiatic/tropical-type carrots because of huge differences in plant stature, root, and leaf traits among Asiatic/tropical-type and Western/temperate-type carrots. He briefly explained the possibility of secondary domestication event due to the natural hybridization between yellow Asiatic/tropical carrots, cultivated white-rooted accessions of wild carrot (*D. carota* L. ssp. *carota*), and adjacent wild populations of *D. carota* subsp (Heywood, 1983). Yet, Vavilov suggested, Asia Minor and the inner Asiatic regions are the centers of origin for cultivated carrots (Vavilov *et al.*, 1992). The molecular studies on carrot domestication also support the cultivated carrot origin in Central-Asia and focus on genomic region contributions during the domestication and evolution, clearly accepting the fact that Western/temperate cultivated carrots may have developed from Asiatic/tropical-type carrots (Iorizzo *et al.*, 2013; Rong *et al.*, 2014).

Transcriptome investigations on storage roots of cultivated carrots, wild carrots, and other wild *D. carota* subspecies accept the hypothesis that Western/temperate-type orange carrot may have originated from Asiatic/tropical-type of carrots. However,

introgressions from wild carrots may also have played a major role in the process (Rong *et al.*, 2014). Wild and cultivated carrots are genetically highly distinguished from one another still; both the groups of carrots retain high and similar levels of genetic diversity. In brief, the alleles might vary amongst wild and cultivated groups resulting in genetic differentiation however, the total number of alleles and level of heterozygosity remain high in both groups (Rong *et al.*, 2014).

Most of the molecular genetics studies have clarified questions relating to carrot evolution and domestication and confirmed that domesticated carrots were derived from wild populations of Central Asian *D. carota* (Iorizzo *et al.*, 2013). Genomics and Transcriptomics studies similarly revealed that the cultivated germplasm can be divided into two distinct groups: the Asiatic/tropical-type and the Western/temperate-type gene pools (Baranski *et al.*, 2012; Iorizzo *et al.*, 2013; Grzebelus *et al.*, 2014), in harmony with the previous hypothesis on carrot evolution by Small (1978), based on the morphological interpretations.

Western/temperate-type and Asiatic/tropical-type carrots demonstrate strong population structure (Shim and Jorgensen, 2000; Bradeen *et al.*, 2002; Rong *et al.*, 2014; Chaitra *et al.*, 2020) at the genome level and also among Asiatic/tropical (Central and Eastern Asia) and Western/temperate (American and European) geographies (Clotault *et al.*, 2010; Baranski *et al.*, 2012; Soufflet Freslon *et al.*, 2013; Maksylewicz and Baranski, 2013; Grzebelus *et al.*, 2014). However, significant overlap in structure between wild and cultivated carrots is often observed because of continuous gene flow between wild and domesticated samples from the Asiatic /tropical-type group, this may be recognized as either recent admixture or domesticated carrots sharing several same alleles as wild carrots from the region. The degree of diversity is high in wild and Asiatic/tropical-type germplasm than Western/temperate-type carrot, which possesses the least amount of diversity (Iorizzo *et al.*, 2013, 2016; Ellison *et al.*, 2018; Chaitra *et al.*, 2020).

Even though most of the genetic elucidations at present support a central Asian center of domestication for carrot, the key mechanisms involved in domestication that contribute to phenotypic changes associated with domestication has to be thoroughly studied.

The carotenoid level is found to be doubled due to carrot breeding in six decades (Simon 1990). Therefore, continuous efforts were made for understanding the mechanisms involved in higher carotenoid accumulation and regulation. Allelic variation at two particular genes, Y and Y2, provides explanations for the distinguishing color and carotenoid accumulation detected in orange, yellow, and white carrot roots (Buishand and Gabelman, 1979). Though, Y or Y2 did not map nearly for carotenoid biosynthesis genes in carrot (Just *et al.*, 2007).

Higher consumer preference and significance for orange carrot increased the understanding prospective of carotenoid accumulation outside the biosynthetic pathway to know the regulatory and other modifying genes for clear details. Iorizzo *et al.* (2016) utilized two mapping populations and the newly assembled high-quality reference carrot genome to recognize a candidate gene outside of the carotenoid biosynthetic pathway viz. The Y gene, DCAR_032551, regulates photosystem development and conditions a portion of carotenoid presence in carrot roots.

The Orange (Or) gene explains the higher levels of carotenoids in cauliflower (Li *et al.*, 2001). The Or gene is destined for the biogenesis of chromoplasts, where carotenoids are stored furthermore post-transcriptional regulation of Phytoene Synthase (PSY), an enzyme necessary for carotenoid biosynthesis (Lu *et al.*, 2006; Yuan *et al.*, 2015; Zhou *et al.*, 2015). Mutations in the Or gene are found to be concomitant with the occurrence of carotenoids in non-leaf tissue as a result of the differentiation of non-colored plastids into chromoplasts in *Arabidopsis*, cauliflower, and sweet potato (Yuan *et al.*, 2015). Likewise, large amounts of β -carotene in the form of crystals in carrot is notably similar to that found in the cauliflower Or mutant as revealed by Maass *et al.* (2009).

Ellison *et al.* (2018) deliberated carrot root pigmentation by utilizing 674 globally distributed domesticated and wild carrot accessions to conduct a genome-wide association study (GWAS). Which enabled the first identification of the 'Or' gene that plays a major role in carotenoid presence in carrots. Mutations in the 'Or' gene are accompanying the increased chromoplast formation, thus allowing more storage competency for carotenoids (Yuan *et al.*, 2015).

The sequence variations at the carrot 'Or' locus have shown non-synonymous mutation co-segregating with carotenoid content, a significant GBS SNP is linked to the Or gene. The wild-type allele (A) seems to be almost fixed in all wild STRUCTURE groups as well as the Asiatic/tropical-type domesticated non-orange samples. The domestication allele (T) was found at high frequency in the Asiatic/tropical domesticated orange samples as well as Western/temperate domesticated non-orange samples and is fixed in all Western/temperate domesticated orange samples. It was observed that domestication genotype (TT) be absent in all wild samples and the wild genotype (AA) be absent in orange samples (Ellison *et al.*, 2018). This mutation was absent in all wild carrot samples and nearly fixed in all orange domesticated samples. The Or domestication allele appears to have been selected after the initial domestication of yellow carrots in the east, near the proposed center of domestication in Central Asia (Ellison *et al.*, 2018).

Further to understand the genomic regions related to domestication events, Iorizzo *et al.* (2016) examined genome-wide F_{ST} between wild and domesticated Asiatic/tropical accessions and found local differentiation signals on chromosomes 2, 5, 6, 7, and 8. Peaks on chromosomes 2, 5, and 7 overlapped with previously mapped domestication QTL for root thickness (cult) (Macko-Podgórní *et al.*, 2017) and carotenoid content, (Y and Y2) (Cavagnaro *et al.*, 2011; Just *et al.*, 2009). Ellison *et al.* (2017) found a drastic decrease in nucleotide diversity in the fine-mapped Y2 region in orange cultivated accessions.

Macko-Podgórní *et al.* (2017) detected a candidate domestication syndrome gene, DcAHLc1, with three non-synonymous single nucleotide polymorphisms and one *indel* that analytically discriminates wild and cultivated accessions. The gene is closer to the AT-hook motif nuclear localized (AHL) family of regulatory genes in plants particularly operational for the regulation of organ development, including root tissue patterning. AHL genes and AHL proteins interaction with other proteins require intracellular protein movement. Researchers get an idea of DcAHLc1 involvement in the development of the carrot storage root, as the localization of the gene overlapped with one of the QTL for root thickening.

The first comparative transcriptome investigation between wild and cultivated carrot roots at multiple developmental stages by Machaj *et al.* (2018) described that transcription factors and genes encoding proteins involved in post-translational modifications were typically upregulated, while those involved in redox signaling was mostly downregulated. Similarly, genes encoding proteins regulating cell cycle, cell divisions, vascular tissue development, water transport, and sugar metabolism were enriched in the upregulated clusters of cultivated carrot.

The root is referred to as the “hidden half” of the plant, so a strong and well-established root system plays an important role in plant performance and yield in the case of crop plants (Lynch 2011; Paez-Garcia *et al.*, 2015). Thus, for developing plants suitable for growing in different agro-ecological conditions, it's very necessary to study the molecular mechanisms underlying the expression of genes and other regulatory factors. Therefore, transcriptomics of two contrasting domesticated cultivars of carrot provides comprehensive emphasis on the set of genes expressed at a time from multiple pathways which would be utilized for suitable crop improvement programs shortly.

2.6 Floral Development and Reproductive Biology

Carrot plants are characterized as temperate and late flowering (Western) with a tendency to shift from vegetative to reproductive phase and induce floral development only after exposure to cold temperature *viz.*, vernalization is mandatory for floral initiation in temperate carrot cultivars. Whereas carrots developed for sub-tropical or tropical climates are early flowering plants, naturally requiring negligible and almost nil subjection to cold temperature for floral induction (Simon *et al.*, 2008).

Heritable genetic mechanisms are involved in reproduction biologies such as flowering time, flower development, floral pattern, pollen fertility, and male sterility, furthermore seed set is crucial for carrot breeding programs. So, the major phases of flower development and pivotal genes and innate mechanism of flowering in carrots are discussed below.

The carrot being a biennial crop, its life cycle consists of two phases *viz.*, vegetative and reproductive/generative phases. Since the seed is the propagating material and the root is the edible part, both phases are considered to be important for

farmers and breeders. The generative phase contains various ontogenetic stages. In the preliminary stages, the plant prepares for flowering with the stimulation of the shoot meristem for stalk elongation (Junttila, 1991). Hence, the shoot apical meristem (SAM) is predominant and is concerned with most of the organogenic phenomena usually accompanying shoot morphogenesis (Wang *et al.*, 2019). While, the sub-apical meristem is the most important site of cell division and elongation furnishing stem extension, eventually transformed into an inflorescence meristem, and the complex process of flower development. Then, floral meristem identity genes regulate the formation of the flower. Once the flowering initiates, floral genes which govern the whorl formation in flowers are expressed, the structures of the whorls and their occurrence at the right position are regulated by homeotic genes (Wang *et al.*, 2019).

Accurate control of the different steps of flower development is achieved by a gene regulatory network (GRN) that is composed of interacting genes and their protein products. Most of the genes in the floral GRN belong to highly conserved gene families encoding transcription factors (TFs), such as the MADS-domain, LFY-like, and AP2-like proteins (Theissen and Melzer 2013). After pollination, fertilization, and seed development, the reproductive cycle ends with the senescence and decline of the plant (Wang *et al.*, 2019).

2.7 Initiation of Flowering - Juvenility, Vernalization and Photoperiodic Requirements

The stage of growth when carrot plants are not responsive to low temperature is known as juvenility (Atherton *et al.*, 1990; Galmarini and Della Gaspera 1996; Galmarini *et al.*, 1992). Carrot varieties growing in Asia tend to be annuals when grown under long days and shall not necessitate vernalization for flowering. Whereas, cultivars advanced in temperate areas of Europe and North America has remained as biennials (Lacey 1986; Verkaar and Schenkeveld 1984; de Jong *et al.*, 2016; Harrison and Dale 1966; Iorizzo *et al.*, 2013).

Carrot has major developmental phases in their lifespan. Wherein, vegetative phase is the main process of root development in the first year of its lifecycle to store photosynthetic material in its root which is subsequently utilized for its reproductive

growth in the second year (Iorizzo *et al.*, 2013; Que *et al.*, 2019). Once the plant has matured beyond a period of juvenility, flowering is stimulated by exposure to a certain period of low-temperature conditions. After stimulation, flower stalk elongation (bolting), flowering, and seed production are promoted by long-day conditions (Atherton and Basher, 1984).

In carrots, the level of cold treatment is cultivar-dependent. For temperate carrots classified as late-flowering cultivars, they need ≈ 11 to 12 weeks at 5°C to be permanently vernalized (Atherton *et al.*, 1990; Hiller and Kelly, 1979; Rubatzky *et al.*, 1999; Dickson and Peterson, 1958). Development to flowering is more rapid with increasing thermal times of vernalization. Short photoperiods (12hr) or darkness during vernalization results in an earlier and a more promoted effect on flowering than long days (16hr), suggesting that temperate carrots are short-long day plants (Atherton and Basher, 1984). Plants that have been induced flower, but are not yet undergoing stem elongation, can be de-vernalized by several days of high temperatures ($28\text{--}35^{\circ}\text{C}$) (Hiller and Kelly, 1979). While tropical plants classified as early flowering cultivars (Alessandro and Galmarini 2007; DiasTagliacozzo and Valio 1994) require shorter vernalization periods (1–4 weeks) than late-flowering cultivars.

The vegetative phase in carrots naturally ends when plants attain 8-12 leaves and storage roots larger than 4–8 mm in diameter (Atherton *et al.*, 1990; Galmarini and Della Gaspera 1996; Galmarini *et al.*, 1992). The floral initiation in carrot undergoes a morphological shift from comparatively flat apical meristem producing leaves, to a heightened conical meristem capable of producing stem elongation and an inflorescence (Hiller and Kelly, 1979). The seed stalk development in carrot is cultivar dependent; it may perhaps get longer to a height of 60-180 cm, forming a highly branched bristling inflorescence having large nodes and hollow internodes (Galmarini *et al.*, 1992; Hiller and Kelly, 1979).

2.8 Flowering and Pollination

The botanical family Apiaceae (previously Umbelliferae) comprises plants whose distinct feature is, the arrangement of their flowers in umbels (Ajani *et al.*, 2016), henceforth their name. Botanically an umbel is an inflorescence comprising of several

short flower stalks (called pedicels) which outspread from a common point appear somewhat like umbrella ribs (Linke *et al.*, 2019).

Carrot flowers are usually perfect, small and white or occasionally greenish-white or pale yellow. Carrots usually consist of pedicellate flowers wherein the stem attaches a single flower to the inflorescence. The flowers of the Umbelliferae *family* are hermaphrodite, epigynous, and contain petals, sepals, and stamens in a set of five hence called, pentamerous. The stylopodium is a swollen and often strongly expanded base of the style in the Umbelliferae family which serves as a nectary, observed as a special form of a gynoeceum nectary or ovary-roof nectary. This is important to increase the attractiveness of insects (Broussard *et al.*, 2017; Mas *et al.*, 2018). Wild and cultivated carrots appropriately exhibit andromonoecy and ensure hermaphroditic and staminate flowers on the same inflorescence. The proportion of both flower kinds is variable in umbels of a different order and has a habit to produce staminate flowers in subsequent orders. The possibility to originate staminate flowers devoid of ovaries is maximal on higher-order umbels (Braak and Kho, 1958).

Approximately, 95% of the flowers are hermaphroditic on the primary umbel however, staminate flowers are found more toward the center of the umbel (Lamborn and Ollerton 2000). Individual carrot flowers are protandrous dichogamy (Koul *et al.*, 1993) has recognized as a mechanism evolved to reduce geitonogamy (Koul *et al.*, 1989) and facilitate outcrossing. Self-pollination occurs to a certain extent by coinciding in flowering among different umbels but is undesirable except for plant breeding programs. Whereas, in wild plants (Asiatic/tropical-type cultivars) secondary umbels are merely developed after the primary umbel has been pollinated (Westmoreland and Muntan 1996).

The flowering phase initiated after a period of low-temperature exposure involves several physiological and morphological changes in the anatomy of plants. Roots turn into hard fibrous and uneatable even before the first visible inflorescences appear (Ajani *et al.*, 2016). In numerous carrot accessions and wild relatives, the central umbellet can be reduced to one or a few white or dark-red flowers. The dark central flower was believed to be important for insect pollination was systematically disproven (Lamborn and Ollerton 2000).

The flower development begins once the flat apical meristem switch from leaf production to an uplifted, conical meristem that stimulates stem elongation and produces a cluster of flowers. The cluster is a compound umbel, and each umbel contains numerous smaller umbels (umbellets). The flowers are found in umbellets whereby pedicels of each flower radiate from a common point (Linke *et al.*, 2019).

Flower development is centripetal. Flowering first commences on the primary or king umbel towards the terminal of the central main stem; smaller secondary umbels grow from the main branch, and this further branch into third, fourth, and even later-flowering umbels, the system of branching continues for at least four orders with smaller umbels on each successive order (Hawthorn *et al.*, 1962).

The time and duration of flowering differ hugely between plants, within cultivar and seed fields (Hawthorn, 1952). Individual umbels period of flowering is usually for 7 to 10 days and there is generally overlap in flowering on successive umbel (Hawthorn *et al.*, 1962; Hiller and Kelly, 1979; Rubatzky *et al.* 1999). Depending on the seed production area, variety, cultural practices, and environmental conditions, the total crop bloom period is usually 4 to 6 weeks (Hiller and Kelly, 1979; Rubatzky *et al.* 1999). As carrot flowers are typically entomophilous (Koul *et al.*, 1993) secreting nectar from the disk-shaped nectarines situated on the upper surface of the ovaries (Mas *et al.*, 2018) increases both seed yield and seed quality when there was an adequate supply of pollinating insects.

In the following sections, crucial stages of carrot reproductive biology are highlighted, specifically after the commencement of flowering after vernalization, and the formation of the flower transition is described.

Arabidopsis thaliana has been the most widely studied plant model system in the past few decades, thoroughly annotated genome, repositories, databases, and genetic stock centers has helped to understand the evolutionary breadth of plants. Molecular mechanisms involved in flowering time have been extensively reviewed for *Arabidopsis thaliana*, a model long-day plant (Amasino, 2010, Song *et al.*, 2018).

Flowering initiation is controlled by a series of physiological pathways acting in response to environmental cues and endogenous signals (Andres and Coupland, 2012).

The transition from the vegetative phase to reproduction is a crucial developmental switch in the plant biological clock and the timing of this switch is vital for reproductive success. The decision of transition from juvenile to generative phase is genetically driven by the interaction of endogenous factors (e.g., age and developmental state) and environmental cues. Light and temperature (acting through photoperiod and vernalization pathways) are the most reliable seasonal cues (Amasino, 2010, Andres and Coupland, 2012) by which plants anticipate upcoming changes in environmental conditions and regulate flowering time in plants (Shim and Jang 2020).

Molecular insights into the endogenous and environment sensing pathways have suggested that inputs from multiple pathways are integrated into the flowering decision (Amasino and Michaels 2010). The signaling cues perceived in the leaves and shoot apical meristem (SAM) induce flowering (Koornneef *et al.*, 1991).

2.9 Flower induction by the photoperiod pathway

The molecular genetic basis of flowering time variation is well studied in ecological and evolutionary contexts of *Arabidopsis* (Rosas *et al.*, 2014). Plants have adopted intricate mechanisms to understand variations in day length to accurately begin flowering through a seasonal progression which is called photoperiodism (Garner and Allard, 1925). There are three major photoperiod's response categories; length of the night is a controlling factor (Hamner and Bonner 1938; Hamner, 1940). Short-day plants flower when night exceeds a critical length (normally in autumn); long-day plants flower when night falls below critical length (normally in late spring and summer); and day-neutral plants flower once after attaining a particular developmental stage irrespective of day length (Andrés and Coupland, 2012).

2.10 Regulatory network of flowering in long-day biennial carrot

Late flowering mutants isolated from *Arabidopsis* depending on their increased total number of leaves have provided evidence about key regulatory genes in the process of flowering induction in long-day plants (Rédei, 1962; Koornneef *et al.*, 1991). Genes like FLAVIN-BINDING KELCH REPEAT F-BOX1 (FKF1), GIGANTEA (GI), CRYPTOCHROME2 (CRY2), FLOWERING LOCUS E (FE), CONSTANS (CO), and FLOWERING LOCUS T (FT) (Andrés and Coupland, 2012; Song *et al.*, 2015) are

expressed in leaves where photoperiod signal is perceived (Takada and Goto, 2003; An *et al.*, 2004). However, expression patterns of FKF1 and GI are often found to overlap with that of CO and FT in the vascular tissue of leaves (Song *et al.*, 2013).

LD dependent flowering is associated with transcriptional activation of CO, a member of the B-box (BBX) zinc family which contains two N-terminal B-boxes and a C-terminal CONSTANS, CONSTANS-LIKE, TIMING OF CAB EXPRESSION1 (TOC1) (CCT) DNA-binding domain (Strayer *et al.*, 2000; Robson *et al.*, 2001; Khanna *et al.*, 2009; Gangappa and Botto, 2014). Transcription factor CONSTANS (CO) is a key regulatory component of the photoperiodic flowering time in *Arabidopsis* (Putterill *et al.*, 1995). Transcriptional activation of CO is light-dependent wherein, cis-regulatory variations in coding regions are responsible for fluctuations in flowering time (Rosas *et al.*, 2014).

In carrot, global analysis of genes by transcriptome sequencing by Ou *et al.* (2017) identified 45 flowering time-related unigenes involved in the flowering of carrot. Genes were further characterized into five different classes including photoperiod, vernalization, autonomous and gibberellin pathway, and floral integrators. Several homologs of LATE ELONGATED HYPOCOTYL (LHY), CONSTANS-LIKE 2 (COL2), were found to be abundantly expressed in short days than long days. Dct43207 (COL2) and Dct20940 (COL5) displayed contradictory photoperiodic responses: COL2 was down-regulated by light, whereas the COL5 was induced. FLC, SOC1, and GAI exhibited differential expression between wild and cultivated carrots (Ou *et al.*, 2017). In addition, Zhan *et al.*, 2017, characterized FT (FLOWERING LOCUS T)-like genes with the possible significant association in the promotion of flowering in carrots.

Vernalization is a process, in which prolonged exposure to a critical cold temperature induces flowering, but plants need additional seasonal cues or weeks of growth before they begin to flower (Amasino, 2004). At the molecular level, Flowering Locus C (FLC) is a protein that binds to genes that promote flowering and blocks the flower transition phase in plants. A gene recognized as FRIGIDA (FRI) is known to promote FLC expression and repress flowering (Choi *et al.*, 2011). Continued exposure of carrot to cold induces expression of VERNALIZATION INSENSITIVE 3 (VIN3) which interacts with VERNALIZATION2 (VRN2) polycomb-like complex through

chromatin remodeling to reduce FLC expression (Alessandro and Galmarini, 2007). This makes a direction for vegetative to reproductive phase transitions by enhancing the expression of FT (FLOWERING LOCUS T) and SOC1 (SUPPRESSION OF OVEREXPRESSION OF CONSTANS1) (Helliwell *et al.*, 2006; Michaels and Amasino 1999).

In other biennials like Brassica and Raphanus both within the Brassicaceae family, (Bagget and Kean 1989; Ferreira *et al.*, 1995; Osborn *et al.*, 1997; Lan and Paterson 2000; Kole *et al.*, 2001) and in case of Beet (Reeves *et al.*, 2007), FLC-like genes have been found. This is playing a typical role in biennial plant habits. The level of response to cold treatment to break this FLC is cultivar-dependent, late-flowering cultivars; from seed-to-seed crops need approximately 11-12 weeks at 5 °C to be permanently vernalized (Atherton *et al.*, 1990; Hiller and Kelly 1979). Interestingly, virtually all crop species well-known as biennials, equally have early flowering or annual plants which need to be explored. Early flowering plants need 1-4 weeks of shorter vernalization periods (Alessandro and Galmarini 2007; Dias Tagliacozzo and Valio 1994). Carrots adapted to temperate climate are typically believed to be biennials, but a significant difference is observed both within and among wild carrot populations. It is noticed that biennials co-exist with winter annuals, summer annuals, and monocarpic perennials. The main difference among different taxa is considered to be the direction and timing of floral transition (Ajani *et al.*, 2016; Erbar and Leins 2004).

Wohlfeiler *et al.* (2019) derived nine crosses between biennial and annual phenotypes considering the percentage of flowering of plants as adequate discrimination of annual and biennial plants (parental lines, F1 and F2 families) in carrot. The segregation ratios obtained in this study were used to create a genetic model for this trait, having 2 genes (Vrn-A and Vrn-B) with three alleles controlling the vernalization requirement. Annual flowering behavior was found to be strongly dominant for both genes, with the A1 allele having an epistatic effect over Vrn-B. Vrn-A and Vrn-B interact generating different vernalization requirement levels. In addition, evidence of Villeneuve and Latour (2017) specified that epigenetic factors also influence flower initiation in carrots. Even after vernalization, flowering could be blocked with constant low light, short-day photoperiods, or by a few days' exposures to high temperature (28–35 °C) if stem elongation had not yet occurred (Fisher 1956; Ou *et al.*, 2017). Only long

days, followed by vernalization promote flowering (Atherton and Basher 1984) in carrots.

Environment components that contribute to reproduction are a strongly determined fact, annuals are found to be widely spread and more prevalent in resource-rich environments whereas, biennials and perennials are abundant in nutrient-poor environments (Lacey 1986; Verkaar and Schenkeveld 1984).

The proportion of biennials, annuals, and/or perennials in a population differs with latitude and resource quality of their habitat (Lacey 1986, 1988). Annual plants are commonly having higher reproductive rates than biennial plants since annuals reproduce every year instead of two years as in biennials or three or more years in perennials (Mandel and Brunet, 2019). An annual plant higher growth rate tends to expand and have greater gene flow because more plants are likely to flower each year, as a result, more pollen and seeds are produced every year increasing the gene flow potential (Mandel and Brunet, 2019).

Cultivated plants derived from wild carrots (Iorizzo *et al.*, 2013) besides having bidirectional gene flow between wild and cultivated carrots (Wijnheijmer *et al.* 1989; Hauser and Bjørn 2001) both the classes remain genetically differentiated (Grzebelus *et al.*, 2014; Iorizzo *et al.*, 2013; Shim and Jørgensen 2000).

2.11 Carrot nuclear genome

The first whole-genome sequences of an orange inbred line, ‘DC-27’ carrot were *de novo* assembled and analyzed. But the genome assembly was fragmented and not assembled to the chromosome level, which limited its application for comprehensive genetic and genomic analyses. Further, the CDS of putative genes from the genome data set of ‘DC-27’ were computationally predicted and not experimentally proven. So, the putative gene CDS might be accurate or produce alternative splices that cannot be identified (Xu *et al.*, 2014).

A major milestone for carrot research is accomplished when Iorizzo *et al* (2016), published a high-quality chromosome level, genome assembly (473Mb) of a doubled-

haploid orange carrot DH1 and made it publicly available for all researchers across the globe. Subsequently, it was used in multiple experiments of the genetic and genomic studies of carrot to decipher the candidate genes controlling key important traits like beta-carotene (Ellison *et al.*, 2017, 2018) pathway, anthocyanin accumulation (Iorizzo *et al.*, 2019), a terpenoid pathway (Keilwagen *et al.*, 2017). Also, in the identification of candidate genes for root development and domestication (Macko-Podgórní *et al.*, 2017), to study the transcriptome profile associated with flowering, carotenoid accumulation, and root development (Machaj *et al.*, 2018; Oleszkiewicz *et al.*, 2018; Ou *et al.*, 2017; Perrin *et al.*, 2017; Que *et al.*, 2018). Further, to clarify phylogenetic associations within the *Daucus* germplasm (Arbizu *et al.*, 2016a; Mezghani *et al.*, 2018; Chaitra *et al.*, 2020) and enlighten the evolutionary history within the Euasterid II (Kim *et al.*, 2018; Reyes-Chin-Wo *et al.*, 2017) and Euasterid I clades (Dong *et al.*, 2018), it has been explored.

Iorizzo *et al.*, 2016, also deciphered the information about gene content estimations in the carrot genome to understand the proportionate coding region. It was observed that comparison of BES with protein database showed that around 10% of carrot genome represents coding sequences (Cavagnaro *et al.*, 2009) and the number of genes was assessed to be ~25,000, with an occurrence of one gene per 19 kb of genomic DNA.

Compared to the “DC-27” genome assembly, the DH1 genome assembly was the first chromosome-scale assembly, representing a 900 and 7.4-fold increase in terms of scaffold and contig N50, and a 10% increase in genome coverage. Based on flow cytometry, the size of the carrot genome was estimated at ~473 Mb (Arumuganathan and Earle 1991). Early cytogenetic analysis demonstrated that a carrot is a diploid species ($2n = 2x = 18$) with nine pairs of chromosomes (Simon 1984). A chromosome karyotype was established and indicated that four pairs of chromosomes are metacentric and five pairs are sub-metacentric (Iovene *et al.* 2011). Iovene *et al.*, (2011) integrated the cytogenetic and genetic maps and each linkage group was assigned to a specific

chromosome, and six of them were oriented according to the short (north)/long (south) arm of the corresponding chromosome.

2.12 Gene Prediction and Annotation in WGS

Both *de novo* gene prediction and evidence-based predictions were carried out. In *de novo* gene prediction, a set of genes from *Arabidopsis thaliana* and *Solanum Lycopersicum* were utilized for identifying gene function. The evidence-based prediction was crucially based on protein-based homology searches from five closely related or model species and carrot transcriptome data. The transcriptome sequence comprised 58,751 carrot ESTs (Iorizzo *et al.*, 2011) and >500 M Illumina reads from 20 libraries, representing multiple tissue and treatments. All gene models produced by *de novo* prediction, protein-homology searches, and prediction and transcript-based evidence were integrated using GLEAN (Elsik *et al.*, 2007). This approach yielded 32,113 non-redundant gene models. The majority of the gene predictions (98.7%) had cDNA-EST expression evidence, which demonstrated the accuracy of gene prediction.

SwissProt and TrEMBL databases were used for identifying putative functional genes, predicted genes had homology with every corresponding gene identified. Important motifs and domains of genes were determined by InterProScan (Zdobnov and Apweiler 2001). Gene models were aligned against the KEGG database (Kanehisa *et al.*, 2002; Kanehisa *et al.*, 2004) to annotate genes involved in biosynthetic pathways, proteins. Following this approach, about 89% of the genes ensure either known homologs or can be functionally classified. Gene models with no match in these databases were labeled “hypothetical proteins” (Zdobnov and Apweiler, 2001).

2.13 Carrot genome sequencing and assembly

An orange, doubled-haploid, Nantes-type carrot (DH1) was used for genome sequencing by Iorizzo *et al.*, 2016. They used BAC end sequences and a newly developed linkage map with 2,075 markers to correct 135 scaffolds with one or more chimeric regions. The resulting assembly spans 421.5 Mb and contains 4,907 scaffolds (N50 of 12.7 Mb) (Table 1), accounting for ~90% of the estimated genome size (473 Mb). The scaffold N50 of 31.2 kb is similar to those of other high-quality genome assemblies such as potato and pepper. About 86% (362 Mb) of the assembled genome is

included in only 60 super scaffolds anchored to the nine pseudomolecules. The longest super scaffold spans 30.2 Mb, 85% of chromosome 4. More details of the assembled genome are presented in Table 1.

Table 1: Statistics of the “DC-27” and DH1 carrot genomes and gene predictions

Assembly feature	“DC-27”		DH1	
	Number	Size	Number	Size
Assembled sequences	1,85,376	371.6 Mb (78%) ^a	4826	421.5 Mb (88%) ^a
Scaffolds/superscaffolds	1,68,741	371.3 Mb	3468	419.5 Mb
N50 scaffolds/superscaffolds	0.04 Mb	36.6 Mb		
Longest scaffold/super scaffold	0.19 Mb	30.2 Mb		
Contigs	1,85,376	348.8 Mb	30,938	386.8 Mb
N50 Contigs	4.2 kb	31.2 kb		
Anchored sequences	NA [*]	NA	60	361.1 Mb
Anchored and oriented sequences	NA	NA	50	353 Mb
GC content	34.4	34.8		
Fraction Ns	22.7 Mb (6.1%)	34.7 Mb (8.2%)		
Genome annotation				
Total repetitive sequences	NA	NA	193.7 Mb	
Gene models	31,891	–	32,113	108.2 Mb
Genes in pseudomolecules	NA	NA	30,824 (96.0%)	
Noncoding RNAs	1559	151.5 kb	1386	188.9 kb

^{*}NA data not available

Courtesy: Iorizzo et al., 2016

The percentage is calculated considering the estimated carrot genome size of 473 Mb (Arumuganathan and Earle 1991).

2.14 Comparative Analysis and Genome Evolution

Reconstruction of carrot chromosomal blocks descending from the seven ancestral core eudicot chromosomes indicated that the carrot genome has gone through at least 60 chromosome fusions or translocations, possibly due to multiple WGD events. The two lineage-specific WGDs were evident from the distribution of the 4DTV (Fourfold synonymous third-codon transversion rates) of carrot paleohexaploid paralogous genes, whereas genes from the shared eudicot $\square$ WGT were largely lost, likely owing to extensive genome fractionation. Comparative analysis with grape, tomato, and coffee combined with depth analysis of duplicated blocks harboring paralogous genes under the Dc-a 4DTV and Dc-b peaks indicated that a WGT (Dc-b) followed by a WGD (Dc-a) contributed to the diversification of the 9 carrot chromosomes from the 21-chromosome intermediate ancestor (Iorizzo *et al.*, 2016).

Thirteen plant genomes were utilized for elucidation of phylogenomic associations. Results showed that carrot separated from grape*113 million years ago (Mya), diverged from kiwifruit*101 Mya, and diverged from potato and tomato*90.5 Mya, confirming the previous estimates which dated the Asterid crown group to the Early Cretaceous and its radiation in the Late– Early Cretaceous (Bremer 2009). Further, the divergence between carrot and lettuce, both members of the Euasterid II clade, likely occurred*72 Mya. Recently, a comparative analysis between carrot and *Panax ginseng* indicated that ginseng diverged from carrot*50 Mya (Kim *et al.*, 2018) were released, which will provide new insights into the evolution and speciation of the Euasterid II crops.

By using collinear gene block analyses, Wang *et al.*, 2020 inferred that the Dc- β and Dc- α polyploidization events occurred ~ 77 –87 Mya and ~ 46 –52 Mya, respectively. The occurrence time of Dc- β was seemingly near the divergence time of carrot and lettuce, which, according to a previous report, had taken ~ 72 and 93 Mya, respectively. With collinear ortholog analyses, Wang *et al.*, 2020 estimated that the carrot-lettuce divergence occurred 98–111 Mya, indicating that carrot and lettuce do not share the tetraploidization events. In addition, the homologous dot plot of carrot and lettuce showed that the ratio of homologous regions in the two genomes was 4:3, meaning a whole-genome triplication in the lettuce lineage occurred. In summary,

Wang *et al.*, 2020 demonstrate that two tetraploidization events are specific to Apiales and may have led to the formation of the plant lineage.

2.15 Transcriptomics studies in carrot (*Daucus carota* L.)

RNA or protein sequences provide knowledge about gene structure and their expression profile (Bostan and Chiusano 2015). The development of a diverse collection of RNA sequences representing multiple tissues, developmental stages, growing conditions such as biotic or abiotic stress, and genetic backgrounds is a key resource to catalog genes that are expressed and enable the study of genes that are involved in economically important phenotypes.

Early RNA sequencing data for plants included expressed sequence tags (ESTs) which were obtained from one-shot sequencing of a cloned cDNA. The cDNAs used for EST generation were typically individual clones from a cDNA library that were sequenced using Sanger technology. The resulting sequence was a relatively low-quality fragment, and its length was limited to approximately 500–800 nucleotides. ESTs have been extensively used in plant genomics to study gene structure and their expression. However, the techniques used to generate ESTs were low throughput and not cost-effective, compared to next-generation sequencing (NGS) technologies. The advent of NGS has provided a cost-effective strategy to sequence the whole transcriptome without cloning, enabling the development of diverse sets of transcriptome data. This technological advancement represents the foundation to characterize the gene repertoire of several crop genomes (Bostan *et al.*, 2019)

The RNA was extracted from the root and leaves of inbred lines B6274 and B7262 as well as the pool of F4 B493xQAL recombinant inbred lines (RILs). The assembled sequences were integrated with the 18,044 ESTs and were used to annotate carrot genes, identify SNPs or insertion/deletion (indel) and develop ESSR, KASPar markers. Bio project PRJEB3356 includes 11 sets of pair-end sequences from four *Daucus* genetic stocks (6 cultivated carrots and 5 wild carrots) sequenced by Illumina Genome Analyzer IIx (Rong *et al.*, 2014). In this study, 11,369 SNPs were identified. Of these, 622 (out of 1000 tested SNPs) were validated and used to genotype a large set of cultivated carrots, wild carrots, and other wild *Daucus carota* subspecies, primarily of

European origin. Phylogenetic analysis indicated that Asiatic/tropical-type carrots may originate from Western Asia and Western/temperate-type carrots may be selected from Asiatic/tropical-type carrots (Rong *et al.*, 2014). High genetic differentiation exists between cultivated and wild carrots ($F_{st} = 0.295$) showing the strong effects of selection. The reduction in genetic diversity in Western/temperate-type cultivars due to domestication bottleneck/selection may have been offset by introgressions from wild carrots. Differential gene expression patterns between cultivated and wild carrot roots may be a signature of strong selection for favorable cultivation traits (Rong *et al.*, 2014).

As of August 2021, the NCBI SRA includes 139 (~580 Gb) short read (RNA-seq) collections that are stored under 11 bio-projects for different carrot accessions that were published from 2011 to 2021 (Table2). Besides RNA sequencing data, other transcriptome data available for carrot includes mRNA sequence structure that is based on in silico gene model prediction analysis. To date, three-gene model predictions are available for carrot by Xu *et al.* (2014) that includes 78,935 predicted genes.

Table 2: Summary of *Daucus species* short read archives (SRA) data available in NCBI

Bio project accession	Year	SRA Experiments	Data volume, Gbases
PRJNA668894	2020	12	62
PRJNA610539	2020	10	68
PRJNA574412	2019	24	253
PRJNA626692	2020	6	9
PRJNA413468	2017	5	4
PRJNA391808	2017	2	5
PRJNA350691	2016	29	47
PRJNA291977	2015	29	116
PRJEB3356	2013	11	7
PRJNA80035	2011	9	9
PRJNA391410	2017	2	(No information)

Accessed NCBI database on 20.08.2021

Bio-project PRJNA291977 includes 20 sets of pair-end sequences, sequenced by Illumina Hiseq 2000 and representing different DH1 tissues (Iorizzo *et al.*, 2016). The RNA was extracted from seeds, leaves (under physiological/ stressed conditions in three developmental stages), root (physiological and stressed combined), callus, fibrous roots, xylem, phloem, hypocotyl, petiole, whole flowers (two developmental stages), and bracts and buds. This study provided a comprehensive transcriptome analysis for the carrot genome. These sequences were used for gene model prediction and assembly quality verification (Iorizzo *et al.*, 2016).

Ma *et al.* (2018) studied Transcriptome profiling of genes involved in carotenoid biosynthesis and accumulation between leaf and root of carrot (*Daucus carota L.*) because α - and β -carotene were found to accumulate in both tissues. They obtained 46,119,008/48,414,508 raw reads and 45,394,846/47,887,648 clean reads from the carrot leaf and root, respectively. After generating assembly by Trinity 56,267 and 62,427 leaves and root unigenes were obtained. In addition, a total of 18,354 DEGs were found between the carrot leaf and root unigenes, and 99 of these DEGs were found to be involved in carotenoid biosynthesis as revealed by integrated function annotation. In the carotenoid pathway DEGs, DcPSY1, DcZ-ISO, DcCISO2, DcLBCY, DcLECY, DcZEP1, DcZEP2, DcVDE1, DcVDE2, DcNSY1, DcNSY2, DcA8H-CYP707A1.2, DcAAO3a, DcCCD4, and DcMAX1 were expressed dramatically in the carrot leaf compared with in the root. Most of these DEGs were up-regulated in the carrot leaf compared with those in the root. The expression of DEGs may be related to the higher carotenoid pathway flux in the carrot leaf than in the root.

Bio-project PRJNA350691 includes 29 sets of pair-end sequences, sequenced by Illumina Hiseq 2000 sequencer. The plant material was extracted from the root tissue of plants from an F4 population that was derived from a cross between USDA carrot inbred line B493, an orange-colored root line, and Queen Anne's Lace (QAL), a wild-type white-rooted carrot from the USA (Ellison *et al.*, 2017). The sequences were used for quantitative transcriptome analysis to study the gene expression profile associated with the Y2 gene, which conditions β -carotene accumulation in carrots.

Ellison *et al.*, 2017 discovered a single large-effect Quantitative Trait Locus (QTL) on the distal arm of chromosome 7 that overlapped with the previously identified

β -carotene accumulation QTL, Y2. Fine mapping efforts reduced the genomic region of interest to 650 kb including 72 genes. Transcriptome analysis within this fine mapped region identified four genes differentially expressed at two developmental time points, and 13 genes differentially expressed at the one-time point. These differentially expressed genes included transcription factors and genes involved in light signaling and carotenoid flux, including a member of the Di19 gene family involved in *Arabidopsis* photomorphogenesis, and a homolog of the bHLH36 transcription factor involved in maize carotenoid metabolism.

Bio-project PRJNA482951 includes 18 pair-end sets sequenced by Illumina HiSeq 4000 sequencer (Machaj *et al.*, 2018). The RNA was extracted from root tissue collected from a cultivated orange-rooted breeding line, 2874B, and a wild accession of *D. carota* subsp. *commutans*, JKI-W232/07. Root tissue was collected at 3 developmental stages, 55 days after sowing, 110 days after sowing (developing roots), and 165 days after sowing (mature roots). Overall, 3285, 4637, and 570 genes were differentially expressed in the cultivated carrot in comparisons made for young plants versus developing roots, young plants versus mature roots, and developing roots versus mature roots, respectively. They were assumed to be of special interest concerning the development of the storage root. Genes encoding proteins regulating cell cycle, involved in cell divisions, development of vascular tissue, water transport, and sugar metabolism were enriched in the upregulated clusters. Genes encoding components of photosystem I and II, together with genes involved in carotenoid biosynthesis, were upregulated in the cultivated roots, as opposed to the wild roots; however, they were largely downregulated in the mature storage root, as compared with the young and developing root.

Bio-project PRJNA484382 included nine sets of single-end sequences sequenced by Illumina HiSeq 2500 sequencer (Iorizzo *et al.*, 2018). For this experiment, RNA was extracted from root and petiole tissue representing three genetic backgrounds of cultivated carrots with different pigmentations (purple and green petioles, purple and orange roots). The data was used to study the regulatory mechanisms controlling anthocyanin accumulation in carrot root and petioles. Fine mapping was performed in four carrot populations with different genetic backgrounds and patterns of pigmentation. The regions controlling root and petiole pigmentation in chromosome 3 were delimited

to 541 and 535 kb, respectively. Genome-wide prediction of transcription factor families known to regulate the anthocyanin biosynthetic pathway coupled with orthologous and phylogenetic analyses enabled the identification of a cluster of six MYB transcription factors, denominated DcMYB6 to DcMYB11, associated with the regulation of anthocyanin biosynthesis. DcMYB11 is a candidate gene for petiole pigmentation in all the purple carrot sources in this study.

Ou *et al.*, 2017 performed RNA-seq to identify the flowering time genes, and digital gene expression (DGE) analysis to examine their expression levels. The revealed bolting behavior of carrots was influenced by low temperature, illumination intensity, and photoperiod. A total of 45 flowering time-related unigenes were identified, which were classified into five categories including photoperiod, vernalization, autonomous and gibberellin pathway, and floral integrators. Homologs of LATE ELONGATED HYPOCOTYL (LHY) and CONSTANS-LIKE 2 (COL2) were more highly expressed under short-day conditions than under long-day conditions. Homologs of COL2, CONSTANS-LIKE 5 (COL5), SUPPRESSION OF OVEREXPRESSION OF CONSTANS 1 (SOC1), FLOWERING LOCUS C (FLC), and GIBBERELLIC ACID INSENSITIVE (GAI) were differentially expressed between ‘Songzi’ and ‘Amsterdam forcing’. The homolog of COL2 (Dct43207) was repressed by light, but that of COL5 (Dct20940) was induced. A preliminary model of the genetic network controlling flowering time was constructed by associating the results of DGE analysis with correlation coefficients between genes.

Liu *et al.*, 2020 used a published carrot genome that could provide an effective tool for systematically retrieving the key integrator genes of GIGANTEA (GI), CONSTANS-LIKE (COL), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), FLOWERING LOCUS T (FT) homologs in the photoperiod pathway. In this study, the bolting time of wild species “Songzi” (Ws) could be regulated by different photoperiods, but the orange cultivar “Amsterdam forcing” (Af) displayed no bolting phenomenon.

According to the carrot genome and previous *de novo* transcriptome, 1 *DcGI*, 15 *DcCOLs*, 2 *DcFTs*, and 3 *DcSOC1s* were identified in the photoperiod pathway. The circadian rhythm peaks of *DcGI*, *DcCOL2*, *DcCOL5a*, and *DcCOL13b* could be

delayed under long days (LDs). The peak value of *DcCOL2* in Af (12.9) was significantly higher than that in Ws (6.8) under short day (SD) conditions, and was reduced under LD conditions (5.0). The peak values of *DcCOL5a* in Ws were constantly higher than those in Af under the photoperiod treatments. The expression levels of *DcFT1* in Ws (463.0) were significantly upregulated under LD conditions compared with those in Af (1.4). These responses of *DcCOL2*, *DcCOL5a*, and *DcFT1* might be related to the different bolting responses of Ws and Af. This study could provide valuable insights into understanding the key integrator genes in the carrot photoperiod pathway (Liu *et al.*, 2020).

Despite the enormous potential to phenotype and characterize germplasm resources for vernalization requirements, such researches are not many. Till now, phenotyping has been hardly done in open fields. Reliable environments are mandatory for effective selection of carrot breeding stocks and thoroughly described regulation of flowering by photoperiod, vernalization and seasonal regulation of flowering by epigenetic changes need to be understood (Linke *et al.*, 2019). After the genetic control of floral induction becomes increasingly clear with the better-comprehended knowledge of reproductive biology, carefully planned breeding programs can be initiated for adequate commercial seed production or improving genetic gain in carrots.

Meng *et al.*, 2020 characterized a significant number of anthocyanins in black carrots which are not only a good source of natural food colorant but can also provide many health benefits to humans. Taproots of different carrot genotypes were used to identify the candidate genes related to anthocyanin synthesis, with particular a focus on R2R3MYB, bHLH transcription factors, and glutathione S-transferase gene (GST). The RNA-Seq analysis showed that DcMYB6 and DcMYB7 had a genotypic dependent expression and they are likely involved in the regulation of anthocyanin biosynthesis. They were specifically upregulated in solid black taproots, including both black phloem and xylem. DcbHLH3 (LOC108204485) was upregulated in all black samples compared with the orange ones. Meng *et al.*, 2020, also found that GST1 (LOC108205254) might be an important anthocyanin transporter, and its upregulated expression increased vacuolar anthocyanin accumulation in black samples.

Chialva *et al.* (2021) used a high throughput stranded RNA-seq to identify and analyze the expression profiles of lncRNAs in phloem and xylem root samples using two genotypes with a strong difference in anthocyanin production. They discovered and annotated 8484 new genes, including 2095 new protein-coding and 6373 non-coding transcripts. Moreover, we identified 639 differentially expressed lncRNAs between the phenotypically contrasted genotypes, including certain only detected in a particular tissue, and identified DcMYB7 key anthocyanin biosynthetic transcription factor suggested how the regulation of this pathway may have evolved between genotypes.

The carrot genome severely lacks good-quality coverage of transcriptomics data from different tissues and developmental stages under normal conditions and during abiotic and biotic stress treatments. To further model and decipher the mechanisms underlying the biological processes in this plant species, similar data from transcriptomics, epigenomics, proteomics, metabolomics as well as its genetic properties are needed (Bostan *et al.*, 2019).

a) RNA-seq:

The technology of sequencing transcribed RNA followed by transcript mapping and quantification is known as RNA-seq (Wang *et al.*, 2009). It has several solid advantages over micro-arrays and other previously developed high throughput gene expression methods. It is the first sequencing method that potentially allows the entire transcriptome to be surveyed in a very quantitative and high-throughput manner.

There are two distinct methods for converting RNA-seq data into transcript sequences: through alignment to assembled genomic sequences or via *de novo* assembly. Since few organisms have their genome completely sequenced and assembled, RNA-seq has become increasingly relied upon for providing evidence for the identification of transcribed genes and the annotation of exon structure (Haas *et al.*, 2013).

RNA-seq enables investigation of the complex aspects of transcriptomics, e.g., the discovery of novel RNA entities, transcript isoforms, allele-specific gene expression, and verification of candidate mutations in RNA-coding regions, identification of DE genes between two conditions such as treated versus non-treated cells and between wild type and mutant strains in agriculture research (Sripathi *et al.*, 2021). The procedure of RNA-sequencing includes taking samples of purified RNA, shearing it, converting it to cDNA, and sequencing on high throughput platforms such as Illumina Genome Analyzer, Applied Biosystems Solid, or Roche 454 Life-sciences sequencing systems. This process can generate millions of short reads (25-700 bp) taken from one end of the cDNA fragments (Haas *et al.*, 2013). Short reads could also be generated from both ends of each cDNA fragment (paired-end reads). After sequencing, the reads are mapped to the reference genome or transcriptome (Haas *et al.*, 2013). The task at this step is to find the genomic location where each short read best matches the reference genome while allowing errors and structural variation. Next, the mapped reads will be assembled into gene-level, exon-level, or transcript level summaries depending on the aim of the study (Chen *et al.*, 2009; Li and Durbin 2009). Later the summarized data are normalized for gene length and total transcript amount in the sample (Mortazavi *et al.*, 2008; Marioni *et al.*, 2008). Then the expression differences are compared with statistical tests, leading to a ranked list of genes with associated p-values, false discovery, and fold changes estimates (discovery of differential expression) (Sultan *et al.*, 2008). Further, biological insight from these lists can be gained by systems biology approaches (Oshlack *et al.*, 2010).

To identify and rank such associations, a wide range of enrichment analysis tools have been developed in recent years. The term enrichment analysis refers to the examination of the list of genes to determine if they are over-represented among any set of certain processes or pathways members. Different enrichment analysis tools such as GSEA, DAVID iDEP, etc address various challenges while functionally analyzing large gene lists. All these methods scientifically estimate the relationships between AGS (Altered Gene Sets) and FGS (Functional Gene Sets), then statistically underlines the

most enriched (either over or underrepresented) biological annotations out of thousands of linked terms and contents (Oshlack *et al.*, 2010; Reimand *et al.*, 2019).

b) RNA-seq data analysis

In general, parallel sequencing RNA-seq produces large amounts of reads (from millions to billions) (Van Verk *et al.*, 2013). The reads are stored in files, typically FASTQ format (Cock *et al.*, 2010) that contains the nucleotide sequences and the quality values for each nucleotide. The quality control and the visualization for the raw reads are calculated by using FastQC (Andrews 2010), whose results include sequence quality, GC or N content, sequence length distribution, duplication level.

To obtain high-quality data, trimming and filtering of raw reads must be done as the first step of the analysis. This involves removing the low-quality nucleotides towards the 3' end and the adapters from the 5' end of raw reads. Cutadapt (Martin, 2011), FASTXToolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and Trimmomatic (Bolger *et al.*, 2014) can be used for trimming.

c) *De novo* assembly and annotated transcriptome

De novo assembly of the transcriptome (using the pre-processed RNA-seq reads) allows analysis in non-model organisms that have neither genome sequences nor transcriptome information (Hornett and Wheat 2012). Even though a reference genome is available, it is still recommended to complement a gene expression study by a *de novo* transcriptome assembly to identify transcripts that have been missed by the genome assembly process or are just not appropriately annotated (Hölzer, and Marz, 2019).

In general, RNA-seq assembly merges short reads into longer contiguous sequences, which are named contigs (Ekblom and Wolf 2014). Recently, several tools for *de novo* assembly of transcriptomes such as SOAPdenovo-Trans (Xie *et al.*, 2014), Oases (Schulz *et al.*, 2012), Trans-ABYSS (Grabherr *et al.*, 2011), and Trinity (Haas *et al.*, 2013) are widely used. At the time of this thesis, Trinity regarding its high

performance and high success rate in assembling complex/high-quality transcriptome has been demonstrated to be most effective in several comparative studies (Celaj *et al.*, 2014; Chopra *et al.*, 2014). After assembling the RNA-seq reads into contigs, the program builds them into independent Bruijn graphs including splicing isoforms (Grabherr *et al.*, 2011).

For a biological study, a convincing annotation of transcripts is necessary before any analyses. In general, the completely assembled transcript library can be aligned to public databases such as Protein Families (Pfam) (Finn *et al.*, 2016) or gene ontology (GO) (Ashburner *et al.*, 2000) databases for annotation purposes. It is useful to detect the proteins that share highly conserved domains across species. To ensure the reliability of identification, all of the determined proteins can be further analyzed by using UniProt (UniProt Consortium 2015), non-redundant RefSeq from NCBI (Pruitt *et al.*, 2007).

d) Mapping, counting, normalization, and differential expression analysis

Another useful application of RNA-seq is to quantify the abundance of gene expression and to do comparative studies among samples. It has become an important alternative approach to microarrays for transcript profiling. In comparison with hybridization-based microarrays, RNA-seq offers several advantages such as reducing cross-hybridization artifacts, without needing to determine the sequence a priori, detection of low-abundance transcripts as well as reducing high background noise in the experiment (Kogenaru *et al.*, 2012).

Mapping and counting: The abundance level of a transcript signifies the number of reads (fragments) mapped to the transcript. Numerous programs support the mapping of RNA-seq reads to reference sequences. For organisms that contain genome and annotation information, package HTSeq (Anders and Huber *et al.*, 2015), featureCounts (Liao *et al.*, 2014), or TopHat (Trapnell *et al.*, 2012) is used. For non-model organism's bowtie (Langmead *et al.*, 2009)/bowtie2 (Langmead and Salzberg 2012), for example, are widely used tools. The number of reads aligned to each gene (contig) is counted.

Inevitably the raw read counts contain artifacts and errors. Therefore, it is necessary to normalize the raw counts to get a convincing gene expression level.

Normalization: There are several popular normalization methods such as RPKM (Mortazavi *et al.*, 2008), TPMs, DESeq (Anders and Huber 2015), and TMM (Robinson and Oshlack 2010). RPKM/FPKM (reads/fragments per kilobase per million mapped reads) is used for reducing feature-length and library-size effects within samples. TPMs (transcripts per million) can be used to convert into RPKM/FPKM, but it is more suitable to reduce the errors between samples (Conesa *et al.*, 2005). DESeq and TMM (Trimmed Mean of M values implemented in package edgeR) (Robinson *et al.*, 2010) are the most cited normalization and estimation programs.

Differential expression analysis: By applying statistical methods, differentially expressed genes (DEG) can be identified by comparing the normalized expression level of transcripts among samples. The RNA-seq data follows a positive and skewed discrete distribution, a Poisson distribution, or a negative binomial distribution as models are chosen to account for the biological and technical variability in DESeq and edgeR (Rau *et al.*, 2015). In this thesis, DESeq and edgeR have been used.

In the context of this work, RNA-seq is mainly used to identify/characterize transcriptomes, classify protein families and quantify the expression of genes. This thesis focuses on the computational analysis of transcriptomes from Asiatic/tropical-type and Western/temperate-type carrot storage root and FPL tissues have been done.

2.16 Validation of RNA-seq results with quantitative real-time PCR (qRT-PCR)

Microarray could analyze thousands of genes at a time, only one gene could be analyzed in the qRT-PCR method. One requirement which was inevitable throughout this evolution was that of an appropriate reference gene to normalize the RNA levels. Microarray and real-time PCR techniques are highly sensitive and can calculate the precise fold change during comparative expression analysis. This precision is highly

dependent on the expression of the reference gene. It is most important that the reference gene exhibited stable expression during various treatments considered for comparison (Joseph *et al.*, 2018).

Comparative gene expression analysis requires one or more reference genes to reflect the amount of RNA or cDNA transcribed under a particular condition or treatment in an experiment. While analyzing two samples of different experimental conditions, or tissues, expression of the reference gene is taken as internal standard (Joseph *et al.*, 2018; Tian *et al.*, 2015).

RNA sequencing (RNA-seq) results are more accurate as this technique provides information about gene expression at the genome-scale level (i.e., quantify the expression level of all genes). While qRT approaches can be used to measure the expression levels of a smaller subset of genes. RNA-seq is expected to be reliable as they are based on a sufficient number of biological replicates(three), the added value of validating them with qPCR for selected genes increases the accuracy of RNASeq data. RNA-seq methods and data analysis approaches are robust enough to not always require validation by qPCR and/or other approaches, although there are situations where this may be of added value (Coenye, 2021).

Meng *et al.*, 2020 studied comprehensive analysis integrating RNA-Seq, gene expression analyses to reveal candidate genes related to anthocyanin biosynthesis in different carrot genotypes and tissues. Researchers strongly suggested that DcMYB6 and DcMYB7, bHLH gene, and GST gene were involved in anthocyanin pigmentation in carrot taproots in this study. RNA samples isolated for the RNA sequencing was also used to perform the qRT-PCR analysis to validate the results of RNA-Seq. Five DEGs were randomly selected for the analysis of qRT-PCR. FPKM value was calculated for all samples, and the Log2 ratio of FPKM was used to reflect the fold changes compared with CT. These results showed that these genes in RNA-seq and qRT-PCR shared similar expression patterns. Taken together, RNA-Seq results were confirmed by the qRT-PCR analysis.

Ma *et al.*, 2018 analyzed the expression profiles of 12 carotenoid biosynthesis genes by qRT-PCR validation to further understand the carotenoid biosynthesis mechanism as revealed by integrated function annotation. This result was consistent with the results from the quantitative real-time PCR analysis of DEG expression profiles. In the carotenoid pathway, DcPSY1 and DcLBCY were highly expressed in the leaf but lowly expressed in the root, contrary to the expression levels of DcPSY2, DcPDS, DcCISO1, and DcBCH1. Moreover, DcZEP1 and DcVDE1 presented significantly high expression patterns in the leaf, and these patterns were similar to the transcriptome expression profiles. The expression levels of DcZDS1 and DcLECY had no significant difference between carrot leaf and root. In addition, DcBCH2 and DcECH showed extremely low expression levels in both leaf and root. These genes were expressed differently between carrot leaf and root, maybe resulting in the different accumulated levels of carotenoid. Most of these DEGs were up-regulated in the carrot leaf compared with those in the root. The expression of DEGs may be related to the higher carotenoid pathway flux in the carrot leaf than in the root. These results will help to further understand the carotenoid biosynthesis in carrots.

Kodama *et al.*, 2018 identified transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. Nine MYB, bHLH, and WD40 genes may function as positive or negative regulators in the carrot anthocyanin biosynthesis pathways. Of those, 11 genes were consistently up-or downregulated in a purple color-specific manner across various calli and cultivar comparisons. The expression of 10 out of these 11 genes was validated using real-time quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Results provide insights into regulatory genes that may be responsible for anthocyanin production in carrots and suggest that future efforts could be directed towards understanding how these candidate genes regulate the anthocyanin biosynthesis pathways in this species.

Xu *et al.*, 2020 studied transcriptional analysis of structural genes involved in cyanidin-based anthocyanins biosynthesis Transgenic ‘Kurodagosun’ carrying

DcMYB113 driven by the CaMV 35S promoter had a purple root and petiole, while transgenic ‘Kurodagosen’ expressing DcMYB113 driven by its promoter had a purple root and nonpurple petiole, suggesting that root-specific expression of DcMYB113 was determined by its promoter. DcMYB113 could activate the expression of DcbHLH3 and structural genes related to anthocyanin biosynthesis. DcUCGXT1 and DcSAT1, which were confirmed to be responsible for anthocyanins glycosylation and acylation, respectively, were also activated by DcMYB113. qRT-PCR indicated that all these genes were significantly upregulated in the two 35S: DcMYB113 transgenic ‘Kurodagosen’ carrying DcMYB113 lines compared with the untransformed ‘Kurodagosen’. Together, these results suggested that DcMYB113 positively regulated the expression of DcbHLH3 and all the tested structural genes in the anthocyanin biosynthesis pathway in carrots.

Tian *et al.*, 2015 used nine candidate reference genes (GAPDH, ACTIN, eIF-4 α , PP2A, SAND, TIP41, UBQ, EF-1 α , and TUB) for the selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. Carrot plants were subjected to abiotic stresses (heat, cold, salt, and drought) and hormone stimuli (gibberellin, salicylic acid, methyl jasmonate, and abscisic acid). The expression profiles of the candidate reference genes were evaluated in three technical and biological replicates. Real-time qPCR data analyses were performed using three commonly used Excel-based applets namely, BestKeeper, geNorm, and NormFinder. All nine reference genes displayed a wide range of quantification cycle (Cq) values across experimental samples, indicating variable expression. ACTIN and TUB were the most stable genes identified among all sample groups, but individual analysis revealed changes in their expression profiles. GAPDH displayed the maximum stability for most of the single stresses.

2.17 Assessment of stability, adaptability, and yield performance by Eberhard-Russel test of stability analysis of several genotypes

Spaldon *et al.*, 2017 studied stability analysis and genotype $\times$ environment interaction of quality traits in tomato (*Solanum lycopersicum* L.). The investigation was conducted during the spring-summer season of 2012- 2013 and 2013-2014 over six

environments to study the stability of 25 genotypes of tomato for quality traits. Genotypes $\times$ environment interaction was significant for lycopene, beta carotene, and ascorbic acid which indicated the differential response of genotype to various environmental conditions and non-significant for the number of locules per fruit, pericarp thickness, total soluble solids, and fruit pH. None of the genotypes was stable for all the traits and stability for one trait was independent of stability for other traits. Genotype Rupali was found stable for the number of locules per plant, while genotypes Karan for pericarp thickness. Genotypes were assessed for their stability of performance over environments by the method described by Eberhart and Russel (1966). The genotypes Kubergeeta, Arka Saurabh, and Arka Meghali were identified as stable genotypes for fruit pH. The genotypes Aditya, Karan, Arka Vikas, and Arka Meghali can be utilized in a further quality breeding program.

Parmar *et al.*, 2016 conducted a study on different stability procedures for the yield of Rice genotypes (*Oryza sativa* L.) Twenty genotypes of the late-maturing group were tested at four locations of Gujarat state along with one check GR-103 during *Kharif*- 2006. In the combined ANOVA, the significance of genotype variance indicated the genetic diversity among genotypes yield. All the three components in Eberhart and Russell model, *i.e.*, environment (Lin.), $G \times E$ (Lin.) and pooled deviations were found highly significant and accounted for 86.8, 7.9, and 5.2 percent of the total variance of environment and GEI. Genotypes NWGR-3006, IET-19140, IET-19123, IET-19147, and IET-19160 were found stable and adaptable to all the environments. Genotype IET-19146 was stable and well adaptable to a favorable environment, whereas NWGR-3213, IET-19132, and GR-103 were found suitable for poor environments with lower yields.

Raj *et al.* (2019) carried out to identify stable Maize hybrids across various environments as the performance of each hybrid tends to vary when grown in different seasons or locations. Twenty-one Maize hybrids and two commercial checks were tested over three locations in India viz., Villupuram, Trivandrum, and Nagercoil. Eberhart and Russell's model of stability analysis was carried out which revealed a significant effect of each environment on the hybrids taken, for all the ten

morphological traits except the number of leaves. The hybrid AU-101 was identified as a stable hybrid with a high mean under less favorable conditions and the hybrid AU-114 was recognized as a stable hybrid under favorable conditions. None of the check hybrids viz., CP-818 and Bioseed-TX369 showed stability in any of the environments. Thus, it emphasized the need for the development of location-specific hybrids or a hybrid that is, stable across environments in maize.

The details of the genes identified till today in carrots for different target traits of interest are presented in Table 3 below.

Table 3: List of genes identified in a carrot for various traits of interest globally as on 2021

Gene symbol (parentheses indicate suggested symbol)	Character description/trait	Reference
Reproductive biology		
Vrn1	Vernalization	Alessandro <i>et al.</i> (2013)
Ms1-Ms-3 Ms-4-Ms-5, ms, a, B, D, E, l, t, Rf1	Nuclear restorers of CMS	Thompson (1961); Banga <i>et al.</i> (1964); Mehring-Lemper (1987); Alessandro <i>et al.</i> (2013)
Gum1-2, Mar1-2, Gad1-2	Novel cytoplasms and sterility	Borner <i>et al.</i> (1995)
STS1–STS6	Petaloid male-sterile and fertile cytoplasm	Nakajima <i>et al.</i> (1999)
Morphology and growth		
Cr	Root cracking	Dickson (1966)
Gls	Glabrous seed stalk	Morelock and Hosfield (1976)
sp 1, sp 2	Spine formation	Nieuwhof and Garritsen (1984)
Phenl	Small, dark green, annual	Schulz <i>et al.</i> (1994)
COLA	Compressed lamina	Budahn <i>et al.</i> (2014)
YEL	Yellow leaf	Nothnagel <i>et al.</i> (2005)
cult	Root thickening	Macko-Podgórní <i>et al.</i> (2017)
5, 4, and 3 QTL	Shoot height, biomass, area	Turner <i>et al.</i> (2018)
1, 5, and 3 QTL	Petiole number, width, and length	
6, 2, and 2 QTL	Root length, biomass, and area	
Disease and pest resistance		
Ce	Cercospora leaf spot	Angell and Gabelman (1968)
Eh	Powdery mildew	Bonnet (1983)
3 QTL	Alternaria leaf blight	Le Clerc <i>et al.</i> (2009)
11 QTL	Alternaria leaf blight	Le Clerc <i>et al.</i> (2015)
Mj-1	<i>M. javanica</i> root-knot nematodes	Simon <i>et al.</i> (2000)
Mj-2	<i>M. javanica</i> root-knot nematodes	Ali <i>et al.</i> (2014)
7 QTL	<i>M. incognita</i> root-knot nematodes	Parsons <i>et al.</i> (2015)

Storage root pigments, carbohydrates, and terpenoids		
y	Yellow xylem and phloem	Laferriere and Gabelman (1968); Simon (1996)
y1	Differential orange phloem/xylem	Bradeen and Simon (1998)
y2	Differential orange phloem/xylem	Just <i>et al.</i> (2007, 2009); Yildiz <i>et al.</i> (2013); Ellison <i>et al.</i> (2017)
o, io	Orange xylem	Kust (1970)
a, l	a-Carotene, lycopene	Umiel and Gabelman (1972)
Rp	Reduced carotene content	Goldman and Breitbach (1996)
16 QTL	Carotene content	Santos and Simon (2002)
Or	Carotene content	Ellison <i>et al.</i> (2018)
g	Petiole anthocyanins	Angell and Gabelman (1970)
P1	Root anthocyanins	Cavagnaro <i>et al.</i> (2014)
P2	Node anthocyanins	Simon (1996); Cavagnaro <i>et al.</i> (2014)
Raa1	Acylated anthocyanins	
15 QTL	Anthocyanin content	
30 QTL	Volatile terpenoid content and composition	Keilwagen <i>et al.</i> (2017)
Rs	Reducing sugar	Freeman and Simon (1983); Vivek and Simon (1999); Yau and Simon (2003)

3. MATERIAL AND METHODS

Agriculture is coping with the climate change; however, it requires careful management of resources like soil, water, and exploration of germplasm diversity. Making agriculture sustainable is strategic and is possible only through diverse production systems that make the most efficient use of environmental conditions without damaging the crop system. The climate change impact can be minimized only by the incorporation and implementation of well-designed experiments, which will help to reduce vulnerability, stabilize food production, and better secure livelihoods. The cultivated area of carrot is growing worldwide and its adaptation to different environmental conditions outside the temperate climatic zone would provide health benefits to human populations. This helps in the worldwide supply of carrots for the community of people suffering from malnutrition and vitamin A deficiency disorders (Grzebelus *et al.*, 2014). The present work was carried out in warmer climatic/tropical region of Karnataka, (India), to have a deeper insight into the differentially expressed candidate genes involved in the adaptation of secondary storage root, VC patterning, and flowering biology between two important cultivar types of carrot *viz.*, temperate-type (Western carrots) and tropical-type (Asiatic carrots) by transcriptome profiling and phenotypic evaluation in detail.

Carrot growers and consumers essentially consider the vegetative phase, to be specific, the storage root as an edible vegetable. Meanwhile, the reproductive phase is also a very important area of genetic research for breeding new cultivars and/or seed production. It is mandatory to understand the genetic control and molecular mechanisms involved in both the root developmental phase and reproduction phase; because the successful seed set is vital for developing varieties suitable for different ecological niches. Candidate genes governing root genotype (Que *et al.*, 2019) and flowering (Linke *et al.*, 2019) have yet remained unknown in carrots.

An experiment was also conducted to understand the major climatic factors required for flower induction of temperate adapted Western carrot cultivar under a warmer climate. An attempt was made to elucidate the candidate genes of the flowering pathway between the contrasting carrot cultivars under a controlled phytotron facility.

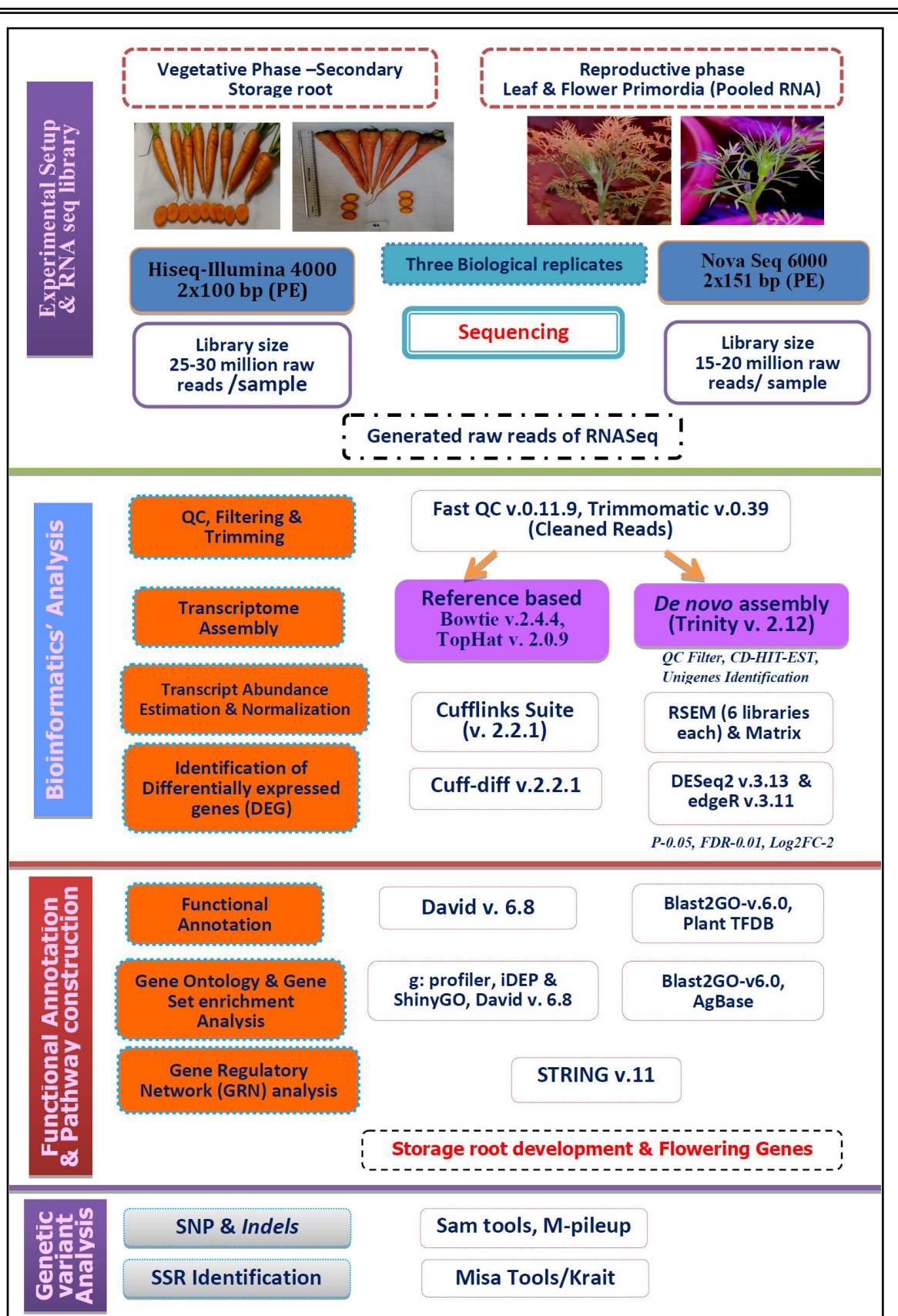
Successfully induced flower primordia and leaf (FPL) in phytotron experiment was subjected to RNA-seq for transcriptomic profiling and the remaining samples were subjected to qPCR validation with five candidate genes related to flowering.

The details of the plant material, methodology of the experiment, statistical and bioinformatics analysis performed in the presented study are presented in the following subheadings.

3.1 Selection of plant material for RNA-seq

Two economically important cultivated types of carrots are globally classified as, Asiatic types- adapted to the tropical region and Western types- adapted to the temperate region were utilized in the present study. One genotype with three replicates, each from two cultivar types viz., namely, UHSBC-23-1 of tropical/Asiatic type and UHSBC-100 of temperate/Western that are being maintained at UHS Bagalkote, were selected for RNA-seq analysis. Meanwhile, Asiatic cultivars are also known as ‘Eastern’ or ‘tropical adapted’ types, and Western cultivars are also known as ‘European’ or ‘temperate adapted’ types (Arbizu *et al.*, 2016a; Spooner *et al.*, 2014; Ellison *et al.*, 2018). Henceforth, the names are sometimes synonymously used throughout the thesis manuscript. Asiatic cultivars are preferably mentioned as ‘tropical-types’ whereas, Western cultivars are cited as ‘temperate types’.

Carrots are propagated by seeds, and the edible part is a secondary storage tap root. Hence, two economically important tissues viz., matured root (vegetative phase) at harvest and floral primordial pooled with apical leaf (transition zone from vegetative to reproductive stage) were subjected to RNAseq in the present work. The details of the experimental workflow used in this work, highlighting plant material selection, RNA-seq computational analysis, annotation, gene ontology, and genetic variant analysis is presented in Flowchart-1. The detailed characters of two contrasting cultivars belonging to temperate and tropical types for root and flower characters are presented in Table 4.



Flow chart 1: RNA-seq experimental work flow and computational analysis

Table 4. Descriptions of two cultivated types utilized for Transcriptome profiling by RNA-seq

Sl. No.	Description	Asiatic cultivar (UHSBC-23-1)	Western cultivar (Commercial Kuruda-UHSBC-145)
1	Scientific Name	(<i>Daucus carota</i> ssp. <i>sativus</i> var. <i>atrorubens</i> Alef.)	(<i>Daucus carota</i> ssp. <i>sativus</i> var. <i>sativus</i>)
2	Geographical diversity	Central and Eastern Asia	Northern Europe; American and European
3	Gene pool	Eastern domesticated orange	Western domesticated orange
4	Ecological Adaptation	Tropical adaptable	Temperate adaptable
5	Root Morphology and leaf characters	Comparatively thicker, shorter, narrow, and conical storage roots. Usually contains pubescent leaves, the presence of lateral branching, and less uniform roots	Smooth and juicy storage roots; cylindrical and ensure fewer pubescent leaves; absence of lateral branching and more uniform roots
6	Pigmentation of the roots	Less outer cortex (phloem); carrot has a strong layer of the inner core (xylem). Clear xylem-phloem patterning. Roots are orange in color with light orange flesh.	A pulpy outer cortex (phloem); carrot appears to be lack of inner core (xylem). No, xylem-phloem patterning Attractive deep orange flesh color root with the self-colored core.
7	Growth habit	Annual growth habit with less or no vernalization requirement. Early flowering-type plants	Biennial growth habit with absolute vernalization requirement. Late- flowering-type plants
8	Flower characters	The dark-red central flowers are observed.	Uniform white umbellets or flowers are observed.
9	End-Use	Consumed raw and for salad purpose Not preferred for canning and poor storage quality.	Consumed raw, sweet, pickles, cake, and juice preparations and also in other food industries. Excellent cultivar for canning and storage.
10	Nutritional	Cortex has meager provitamin A carotenoid content and comparatively high fiber content	Cortex is with high sugar and high provitamin A carotenoid content & less fiber content

3.2 Experimental design

To understand the transcriptome dynamics of root and floral primordia and leaf tissues between two contrasting gene pools of carrot, the experiment was designed as follows:

Number of cultivars: Two

- Eastern/Asiatic/Tropical- UHSBC 23-1
- European/Western/Temperate-UHSBC- 100

Tissues: Two

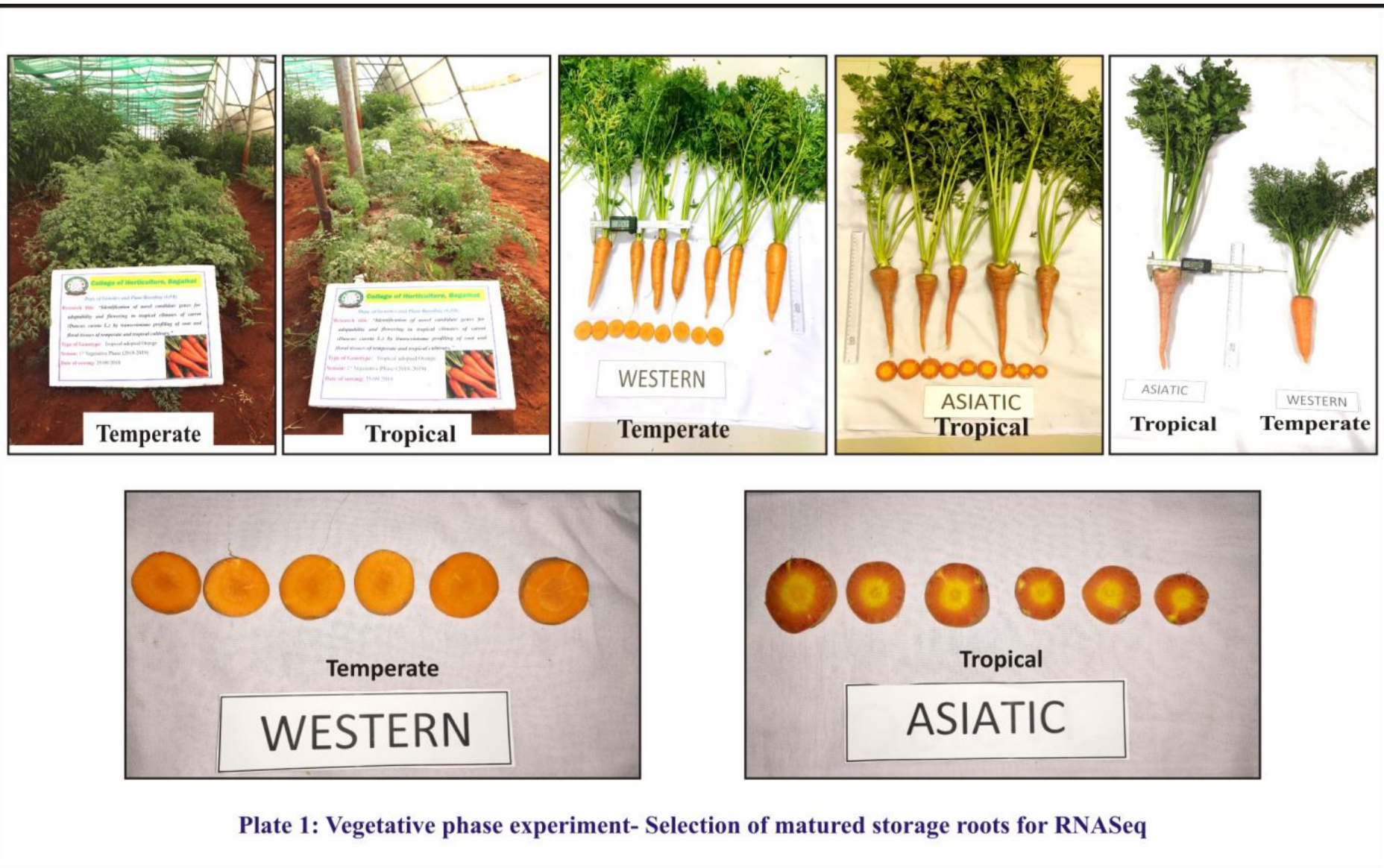
a) Root b) Pooling of floral primordia + leaf tissue of individual plant (FPL)

Number of biological replicates: Three

3.2.1 Vegetative phase Experiment: Selection of roots for RNA-seq

The vegetative phase experiment was conducted from 2019-to 2020. The field experiment was conducted at the newly constructed poly house with fresh and sterile soil, Bagalkot, Karnataka, India during 2019-2020 (Plate 1). Bagalkot is located in the northern region of Karnataka and positioned at 16°12'N, 75°45'E. The average elevation in this area reaches approximately 610 m. The climate is warm and dry throughout the year and rainfall is scarce with an average annual rainfall of 318 mm, belonging to a semi-arid tropical region.

The seeds from the respective cultivars were sown in the replicated trials by bed method (4.0 m × 1.5 m/bed) of sowing in the polyhouse. Freshly obtained sandy loam soil was brought for polyhouse and a month before sowing soil was sterilized with formaldehyde. Previously, there was no sowing of any crop. Carrot root production was followed by all the cultural practices in the vegetative phase of carrot production. Disease-free healthy roots were produced under the polyhouse. 250-300 plants were maintained in each bed with a total of two beds per genotype were maintained. Fresh roots were harvested at 90 days after sowing (DAS) and fully matured and well-developed roots were selected from the pool of harvested roots (Plate 1) at 10.00 am.



Two individual roots from each cultivar constituting one replication were selected. Three biological replicates consisting of six roots from each cultivar were immediately frozen in the liquid nitrogen. The cryopreserved samples were sent for sequencing in the dry ice to *Agrigenomics PVT LTD*, Kochi, Kerala, India on the same day of sample collection.

The shoulder portion (top portion) of a length of 5cm from the remaining roots of both cultivars were separated from the main root and kept for vernalization for 2 weeks (14 days) in the refrigerator at 4⁰ C before sowing for the reproductive phase.

3.2.2 Reproductive phase -induction of flowering and selection floral primordia and leaf (FPL) tissue for RNA-seq

The production of floral primordia or induction of flowering in tropical cultivars (Asiatic cultivars) does not need much effort as they are adapted to a warmer climate where the experiment was conducted (at UHS Bagalkot). However, it was difficult to induce flowering in temperate (Western) cultivars due to the absence of congenial conditions for flowering such as temperature below 12⁰C, vernalization period required during the reproductive phase, long-day conditions, *etc.* in the experimental location. Plate 2 depicts the different stages of floral transition in tropical carrots without the need for vernalization requirement. However, temperate carrot needs mandatory vernalization period (Plate 3).

Hence, to induce the flowering and capture the transition zone of the vegetative to reproductive phase, an experiment was conducted under phytotron with seedlings of previously harvested carrots from both the cultivars. Plants were grown to induce flowering in a phytotron facility funded by KBITS-IT-BT, Govt of Karnataka, being maintained at UHS, Bagalkot during 2020-2021 (Plate 4).

The tissues were collected from the flower primordia (bud size-3.0-5.0 mm) pooled with the leaf tissue to study the flowering related key genes differentially expressed across two cultivars (Plate 5 and 6). Flower primordium is the outcome of an irreversible transition from vegetative to reproductive phase. However, the flowering pathway regulatory genes are primarily perceived in leaves and there is a long-distance

transport of these signals to shoot apical meristem (SAM) where the transition takes place under a specific day length, with a minimum of 7-8 leaves of the plant (Kim, 2020). Hence, flower primordia (3-5mm width) pooled with shoot apical leaves (FPL) were selected from the flower induced plants for RNA-seq and frozen with liquid nitrogen and kept at -80°C until RNA isolation and library preparation for RNA-seq.

Three replicates from each of the cultivars were collected and subjected to RNA-seq. A single plant was considered as one replication with the pooling of an equal quantity of RNA from leaf + floral primordia of that individual.

Two economic tissues from two contrasting cultivars (Temperate/Western and Tropical/Asiatic) of carrots *viz.*, root and flower primordia + leaf were selected respectively at vegetative and reproductive phases of the carrot life cycle. The Hiseq and Nova-seq Illumina sequencing platforms were used, size of the raw reads and the depth of sequencing in both the tissues were different as represented in Table 5.

Objective 1. Transcriptome profiling of root and floral primordia & leaf (FPL) of temperate and tropical adapted genotypes of carrot by RNASeq

3.3 Workflow of Hiseq Illumina and generation of RNA-seq raw reads

3.3.1 Secondary storage root tissue

3.3.1 RNA isolation, quantification & QC report

RNA was isolated from 12 root samples *viz.*, six from tropical, and six from temperate cultivars. Each biological replicate is comprised of two marketable stage roots. The samples were labeled as Tropical_Root_R1, Tropical_Root_R2, Tropical_Root_R3, Temperate_Root_R1, Temperate_Root_R2, Temperate_Root_R3, and the quality of RNA was expressed in RIN (Table 6a) and checked in Bioanalyzer, tape station profile is presented in Fig. 1. Samples with RNA integrity number (RIN) >8.0 were subjected to cDNA synthesis and library construction for RNA-seq.



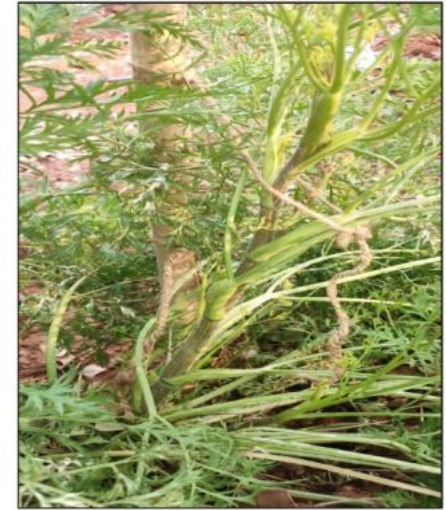
A) Vegetative maturity



B) Increased inter-node extension



C) Increased Stem growth



D) Apical dominance and shoot hardness



E) Increased leaf-growth



F) Vegetative phase change



G) Floral transition



H) Floral primordia induction



I) Umbel development

Plate 2: Vegetative phase change and floral transition in carrot



Asiatic/Tropical carrots

Early flowering- types; No vernalization requirement



Western/Temperate carrots



Late flowering- types; Critical vernalization requirement



Plate 3: Reproductive phase evaluation at field



Phytotron

Plate 4: Phytotron experiment for flowering induction in temperate and tropical cultivars



Temperate cultivars floral primordia + leaf (FPL)

Plate 5: Temperate cultivars floral primordia



Tropical cultivars floral primordia
Plate 6: Tropical cultivars floral primordia

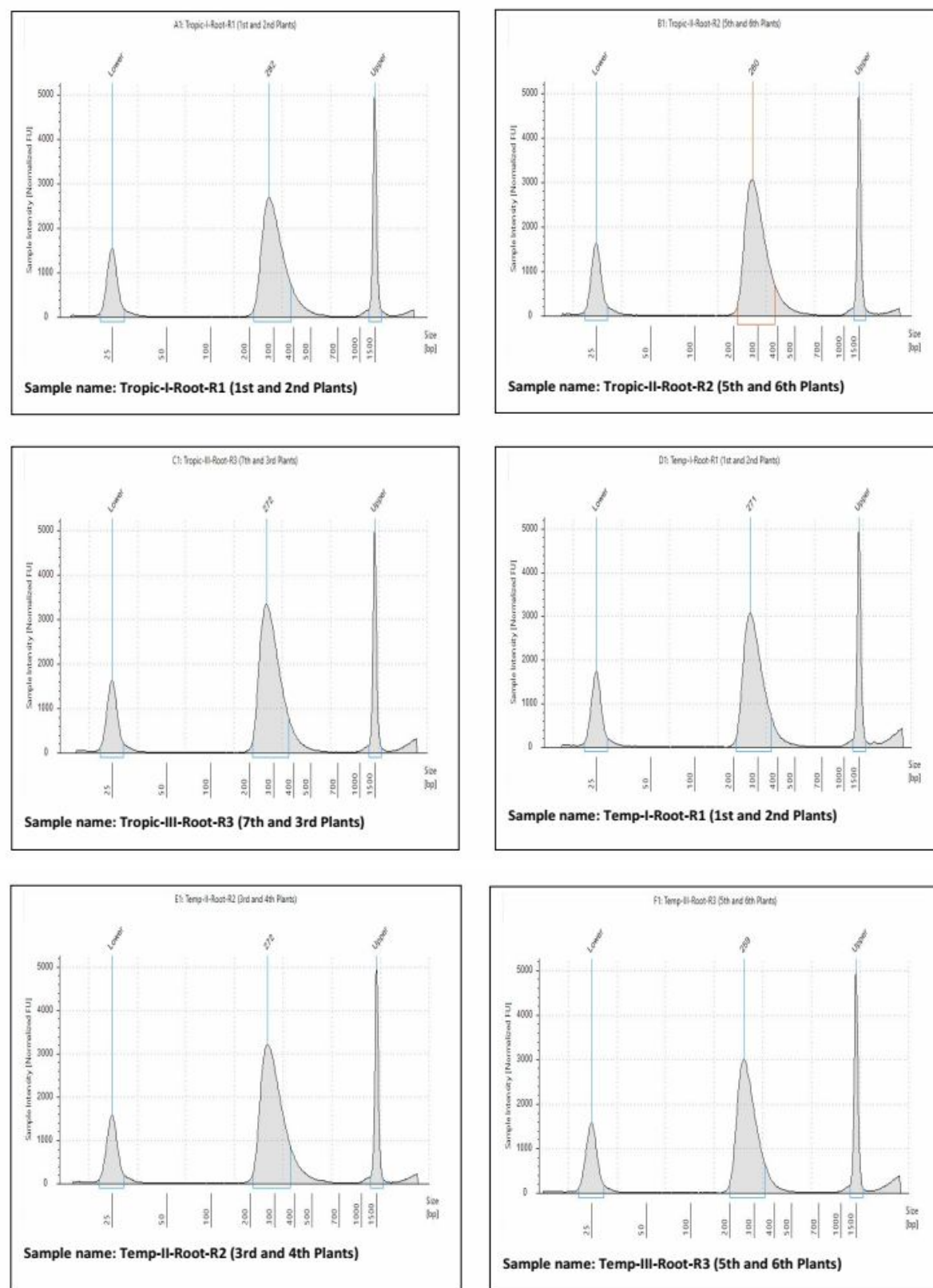


Fig. 1: Tapestation profile Qualitative analysis of RNA of six root samples from Temperate and Tropical cultivars confirmed before RNA-seq by bioanalyzer

3.3.2 Floral primordia and leaf (FPL) tissue

3.3.2a RNA isolation of floral primordia and leaf (FPL) tissues for RNA-seq:

For the samples of the reproductive phase, there were two tissues for pooling to elucidate the flowering pathway in tropical and temperate cultivars. RNA was initially isolated from leaf and floral primordia separately. The quality and quantification of RNA from 12 samples is presented in Table 6b. Three individuals from temperate and tropical cultivars from leaf and floral primordia were labeled as LF_Temperate_R1_Leaf, LF_Temperate_R1_Bud (for Temperate-Replication 1), LF_Temperate_R2_Leaf, LF_Temperate_R2_Bud (for Temperate-Replication 2).

Table 5: Details of experimental plant material and Hiseq Illumina platforms used for Transcriptome profiling (RNA-seq) at vegetative and reproductive phases

Parameters	Vegetative phase Transcriptome profiling	Reproductive phase
Plant Material (cultivars)	1. Temperate/Western cultivar 2. Tropical/Asiatic cultivar	1. Temperate/Western cultivar 2. Tropical/Asiatic cultivar
Tissues	Matured roots at 90 DAS	Pooling of equal quantity of RNA from leaf and floral primordia (3-5 mm bud size)
Growing condition	In newly constructed Poly house with sterilized and fresh sandy loam red soil	Under phytotron chamber in pot culture with sandy loam red soil: sand: vermiculite 1:1:1 pre-sterilized with formaldehyde
Number of replications	Three	Three
Sequencing platform	Hiseq Illumina 4000	NovaSeq 6000
Read Length	Paired end- 2x100 bp	Paired end- 2 × 150 bp Reads
Sequencing depth/library size	25-30 million reads (4-6Gb data/sample)	35-40 million reads (6GB data per sample)

Table 6a: RIN value for root tissue of two cultivars in three replications subjected to RNA-seq

S No	Sample Name	Nanodrop concentration (ng/μl)	Concentration Qubit (ng/μl)	A260/280	RIN Value	Comments
1	Temperate_Root1	934	>1000	2.04	9.20	Can proceed with Library preparation
2	Temperate_Root_R2	686	634	1.97	9.30	Can proceed with Library preparation
3	Temperate_Root_R3	676	628	1.98	9.10	Can proceed with Library preparation
4	Tropical_Root_R1	550	509	1.91	8.80	Can proceed with Library preparation
5	Tropical_Root_R2	508	493	2.02	9.30	Can proceed with Library preparation

Table 6b: RNA integrity number (RIN) value for the leaf and floral primordia/Bud samples of two cultivars in triplicates

S No	Sample Name	Nanodrop (ng/μl)	A260/280	Qubit Concentration (ng/μl)	RIN* Value	Comments
1	LF_Temperate_R1_Leaf	276.70	2.19	278	7.5	Proceeded with Library Prep
2	LF_Temperate_R1_Floral primordia	929.30	2.20	808	7.5	
3	LF_Temperate_R2_Leaf	212.00	2.12	226	6.1	Proceeded with Library Prep
4	LF_Temperate_R2_Floral primordia	800.00	2.10	722	7.2	
5	LF_Temperate_R3_Leaf	190.10	2.14	206	6.4	Proceeded with Library Prep
6	LF_Temperate_R3_Floral primordia	1537.00	2.26	2240	8.0	
7	LF_Tropical_R1_Leaf	570.40	2.07	444	6.5	Proceeded with Library Prep
8	LF_Tropical_R1_Floral primordia	1233.00	2.20	1540	7.9	
9	LF_Tropical_R2_Leaf	936.00	2.11	850	6.3	Proceeded with Library Prep
10	LF_Tropical_R2_Floral primordia	921.00	2.10	610	7.5	
11	LF_Tropical_R3_Leaf	1127.00	2.19	1460	6.0	Proceeded with Library Prep
12	LF_Tropical_R3_Floral primordia	1293.00	2.20	1940	7.4	

*The RNA with RIN value <8.0 were all subjected to RNA purification to improve the quality of cDNA library construction and sequencing

LF_Temperate_R3_Leaf, LF_Temperate_R3_Bud (for Temperate-Replication 3) and LF_Tropical_R1_Leaf, LF_Tropical_R1_Bud (for Tropical-Replication1), LF_Tropical_R2_Leaf, LF_Tropical_R2_Bud (for Tropical-Replication 2), LF_Tropical_R3_Leaf, LF_Tropical_R3_Bud (for Tropical-Replication 3). Temperate indicates Western cultivar and Tropical indicates Asiatic cultivar. The tape station profile of the extracted RNA is presented in Fig. 2.

3.3.2 Generation of RNA-seq raw reads

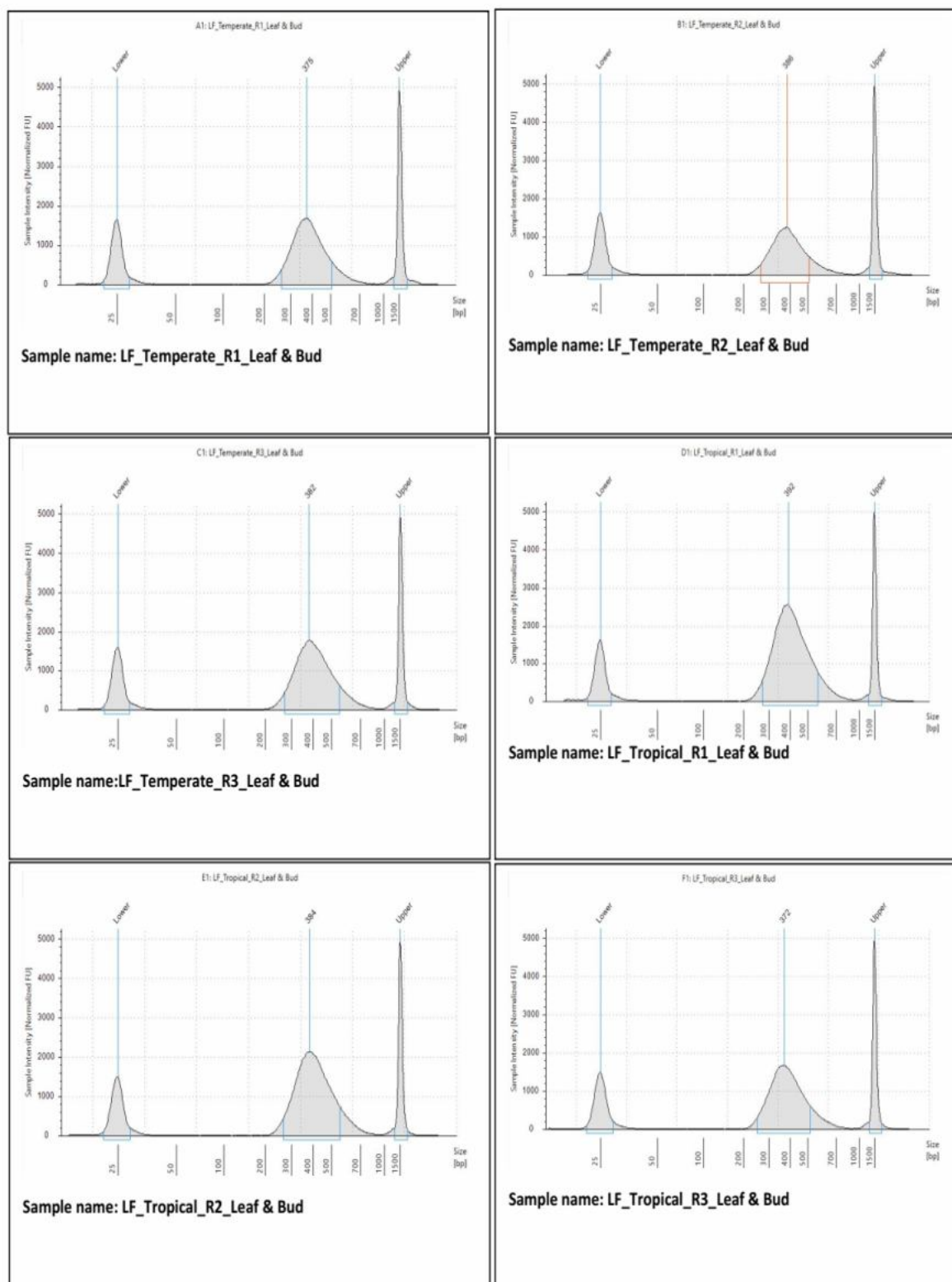
Presently, there are many different high-throughput sequencing platforms, like Roche (454), Applied Biosystems (ABI SOLiD), Ion Torrent, SMART/PacBio, Helicos sequencing system, Nanopore (MinION, PromethION). The different technologies require different experimental protocols and sequencing procedures. The most commonly used ‘Illumina’ sequencing platforms were used in the present study for transcriptome profiling. Illumina currently provides sequencing systems such as MiSeq, HiSeq 2500, HiSeq 3000, HiSeq 4000, HiSeq X Ten, HiSeq X five, NextSeq 550, NovaSeq 6000. In the present study, both HiSeq 4000 and NovaSeq 6000 were used for sequencing carrot root and leaf + floral primordia respectively.

3.3.3 The workflow of Illumina NGS

Principle of Illumina NGS: The Illumina next-generation sequencing (NGS) technique is based on sequencing-by-synthesis (SBS), and reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands

Library construction for Transcriptome Profiling by RNA Seq: ~5-10 ug of total RNA was used to prepare the RNA seq library using the TrueSeq RNA Sample preparation Kits (Illumina).

The libraries were prepared as per the kit protocol. In short, polyA containing mRNA molecules were purified using poly-T oligo attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperatures. The cleaved RNA fragments were used to synthesize first-strand cDNA using reverse transcriptase and random primers followed by second-strand cDNA synthesis using DNA polymerase I and RNase H. These cDNA fragments were then subjected to an end repair process, the addition of a single ‘A’ base, and then



***Bud = Floral primordia**

Fig. 2: Tapestation profile of RNA of leaf and flower bud of Temperate and Tropical cultivars in triplicates confirmed before RNA-seq by bioanalyzer

ligation of the adapters. The products were purified and enriched with PCR to create the final cDNA library. Bioanalyzer plots were used at every step to assess mRNA quality, enrichment success, fragmentation sizes, and final library sizes. The size distribution of the sequencing library was determined by gel electrophoresis. Qubit (Q) was used for measuring the quantity of the library before sequencing. After the libraries were constructed, they were sequenced on HiSeq 4000 to obtain $2 \times 100\text{bp}$, 25-30 million reads/sample (raw reads data) for root tissues. 80% of data of the root sample was Q30 and above. The QC report for 6 samples is presented in Table 7a. For leaf and floral primordia, the Novaseq6000 instrument was used for sequencing that generated 2×151 bp paired-end reads (Table 7b).

Raw reads were received from Agri genomics Pvt Ltd for ‘root samples’ of the vegetative phase during June 2019 (Table 7a) and pooled samples of leaf + floral primordia in a reproductive phase were obtained during September 2021 (Table 7b). The raw data of root tissue and leaf + floral primordia (FPL) were independently analyzed.

3.4 Computational analysis for Transcriptome assembly and Transcript quantification

Bioinformatic analysis and computational challenges of NGS data for both ‘secondary storage root’ and ‘FPL’ tissue RNA-seq data was accomplished by **Dr. Sarvamangala S Cholin. Assistant professor (GPB); PI: PMB laboratory, College of Horticulture, UHS, Bagalkote. Karnataka (India).** **Dr. Akhilesh Bajpai** from *Shodhaka Life Sciences* helped to troubleshoot some of computational encounters. The details of the computational analysis followed for the data generated by RNA-seq for two cultivars and two tissues are discussed below in detail and mentioned briefly in Flowchart 1. The tools used in the pipeline are set with their default parameters with few modifications.

3.4.1 Analysis of raw reads

The standard format in which reads are stored in the FASTQ file. FASTQ file always contains 4 lines per entry and its format is strictly defined so that it can be

Table 7a: Raw reads obtained from RNA-seq by paired-end (PE) sequencing (2x100bp) for root tissues of temperate and tropical cultivars in triplicates

Sl. No.	Sample	Read orientation	Mean read quality (Phred Score)	Number of reads	% GC	% Q < 10	% Q 10–20	% Q 20–30	% Q > 30	Number of bases (MB)	Mean read length (bp)
1	Temperate-R1	R1	39.94	26796143	49.58	0.00	0.69	1.48	97.83	2679.61	100
		R2	37.76	26796143	48.87	0.77	4.71	4.41	90.11	2679.61	100
2	Temperate--R2	R1	39.93	19698951	45.41	0.01	0.69	1.51	97.79	1969.9	100
		R2	37.42	19698951	45.34	0.73	5.57	4.80	88.91	1969.9	100
3	Temperate--R3	R1	40.06	22604096	48.60	0.01	0.58	1.20	98.21	2260.41	100
		R2	38.3	22604096	48.12	0.12	4.22	3.97	91.69	2260.41	100
4	Tropical-R1	R1	39.94	29096070	47.88	0.00	0.69	1.47	97.84	2909.61	100
		R2	37.45	29096070	47.39	0.77	5.4	4.86	88.97	2909.61	100
5	Tropical-R2	R1	39.94	19089333	47.17	0.00	0.7	1.47	97.83	1908.93	100
		R2	37.55	19089333	46.61	0.76	5.17	4.69	89.38	1908.93	100
6	Tropical-R3	R1	39.95	23598359	47.09	0.00	0.68	1.44	97.87	2359.84	100
		R2	37.75	23598359	46.62	0.76	4.82	4.38	90.04	2359.84	100

Table 7b: Raw reads obtained from RNA-seq by paired-end (PE) sequencing (2x151bp) for FPL tissues of temperate and tropical cultivars in triplicates

Sl. No	Sample	Read orientation	Mean read quality (Phred score)	Number of reads	% GC	% Q <10	% Q 10–20	% Q 20–30	% Q > 30	Number of bases (MB)	Mean read length (bp)
1	FPL Temperate R1 Pool	R1	35.76	31726236.00	42.84	0.01	2.77	4.28	92.94	4790.66	151.00
		R2	35.00	31726236.00	42.72	0.01	4.83	6.16	89.01	4790.66	151.00
2	FPL Temperate R2 Pool	R1	36.02	48278351.00	43.79	0.00	2.03	3.74	94.23	7290.03	151.00
		R2	35.42	48278351.00	43.58	0.00	3.65	5.26	91.09	7290.03	151.00
3	FPL Temperate R3 Pool	R1	35.91	44634593.00	42.68	0.00	2.34	4.00	93.66	6739.82	151.00
		R2	35.23	44634593.00	42.64	0.00	4.24	5.52	90.24	6739.82	151.00
4	FPL Tropical R1 Pool	R1	35.97	35856048.00	42.26	0.00	2.19	3.86	93.95	5414.26	151.00
		R2	35.26	35856048.00	42.08	0.00	4.00	5.87	90.13	5414.26	151.00
5	FPL Tropical R2 Pool	R1	36.00	34861066.00	42.30	0.00	2.09	3.80	94.11	5264.02	151.00
		R2	35.36	34861066.00	42.02	0.00	3.69	5.64	90.67	5264.02	151.00
6	FPL Tropical R3 Pool	R1	36.02	37936428.00	42.88	0.00	2.03	3.79	94.17	5728.40	151.00
		R2	35.35	37936428.00	42.66	0.00	3.88	5.36	90.75	5728.40	151.00

always correctly parsed by any software that needs to work with the reads. The file contains the sequences and quality score. The quality score (Q score, or Phred score) is an integer value and it is logarithmically related to the probability of an error in base call (Ref). Formally, quality score is defined as follows:

$$Q = -10 \log_{10} (P)$$

Where, Q = quality score, P = probability of an error ("FASTQ files", 2017). The quality scores in the FASTQ are encoded in ASCII, the encoding is not standardized and depending on the machine it can belong to two groups: ASCII_BASE 33 and ASCII_BASE 64, they can further differ by the value range. FASTQ is the standard output of Illumina, but other formats also exist.

3.4.2 Quality control and processing of raw reads:

The first step before the assembly is to create basic statistics of the reads and analyze their quality. The data received from the sequencing lab can sometimes be poor in quality so checking and running analysis can avoid future problems in assembly, or even reveal that it is completely unusable (Ekblom and Wolf 2014). The most widely used tools for this analysis are FastQC (Andrews, 2010) and MultiQC (Ewels *et al.*, 2016), a quality control tool for high throughput sequence data that was used in the present study for both root and FPL tissue transcriptomes separately.

FastQC runs several tests on the data and reported the basic statistics viz., number of sequences, sequence length, GC content, *etc.* Then performs more complex statistics that can reveal or suggest if there is a problem adapter contamination, plasmid contamination, unbalanced coverage, *etc.* (Andrews, 2010). Raw sequences of six carrot root samples paired-end library of the vegetative phase experiment were checked for quality using the FASTQC (Version 0.11.9) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and analyzed the data including various parameters viz., basic statistics, adapter content, sequence length distribution, sequence duplication level, per base sequence quality and GC content.

Depending on the errors reported by FastQC, an additional step before assembly called pre-processing was carried out. An example of commonly used pre-processing

software that can trim low-quality reads and clipping adapter contaminations is Trimmomatic (Bolger *et al.*, 2014). The adapter was trimmed for each sample individually by applying Trimmomatic v0.36 (Bolger *et al.*, 2014). The parameters with min qual D 28, min len D 50, LEADING: 28, TRAILING:28, SLIDINGWINDOW: 10:28, MINLEN:50 was assigned. After filtering duplicate and low-quality reads, 130618433 high quality cleaned reads were remained out of 140882952 (92.71%). MultiQC (Ewels *et al.*, 2016) was used to generate aggregated reports. Whereas, for FPL tissues, out of total 232922272 generated paire-end reads, 202359387 (86.74%) were used for downstream analysis. The details of reads finally used for assembly/alignment are presented in the results sections.

Further clean reads were used for assembly construction. Two methods of assembly construction were performed for roots *viz.*, reference-based assembly and *de novo* assembly. For leaf/floral primordial data only *de novo* assembly was constructed to explore the novel transcripts for flowering pathway in the transcriptome background of tropical/Asiatic cultivar as the information on this is lacking in the currently available reference genome (Linke *et al.*, 2019). The details of assembly by both methods are described below.

3.4.3 Read alignment/Assembly (Mapping reads to a reference genome or transcriptome)

3.4.3.1 Reference-based assembly (only for roots tissues)

Short sequence reads generated by sequencers after quality check must first be mapped or aligned to a reference transcriptome or genome assembly to discover their true locations (origins) concerning that reference. Genome assembly affords the discovery of novel genes and isoforms, which requires the ability to align, reads across splice junctions. Bowtie 2 (v2.4.4) and TopHat2 (v2.0.9) read mapping algorithms were used for reference-based assembly. TopHat (Trapnell *et al.*, 2009) and Cufflinks v. 0.17.3 (Trapnell *et al.*, 2010) provide a complete RNA-seq workflow; this protocol is adopted by using the reference genome of carrot to identify DEGs across tropical and temperate root transcriptome.

Bowtie is an ultrafast, memory-efficient alignment program for aligning short DNA sequence reads to large reference genomes. For contiguous alignment to a reference, the most well-known aligner is Bowtie. Bowtie indexes and compresses a genome sequence using a technique called a Burrows-Wheeler (BW) transform. The BW transformation allows the index to be held in memory and is faster than other aligning approaches such as spaced seed indexing (Trapnell and Salzberg, 2009). In the present study, Bowtie2 was used to get the index files of the Carrot reference genome fasta file (ASM162521v1-Refseq GCF_001625215.1) downloaded from ncbi.nlm.nih.gov, and TopHat was used for reference-based assembly of the individual carrot sample reads with the help of generated index file and reference genome fasta sequence as explained below.

TopHat (Trapnell *et al.*, 2009) is a fast splice junction mapper for RNA-Seq reads. Spliced aligner, TopHat relies on an approach by which reads are initially aligned contiguously to the genome with Bowtie and the non-aligning reads are set apart. When reads are longer than 75bp they are split into smaller segments and aligned independently. TopHat uses Bowtie as an alignment ‘engine’ and breaks up reads that Bowtie cannot align on its own into smaller pieces called segments. Often, these pieces, when processed independently, will align to the genome. When several of a read's segments align to the genome far apart (e.g., between 100 bp and several hundred kilobases) from one another, TopHat infers that the read spans a splice junction and estimates where that junction's splice sites are. By processing each ‘initially unmappable’ read, TopHat can build up an index of splice sites in the transcriptome on the fly without prior gene or splice site annotations.

The trimmed sequences of the present work from both tropical and temperate root tissues were mapped with TopHat v. 2.0.9 (Trapnell *et al.*, 2009) to the ***Daucus carota* (ASM162521v1)** reference genome. A total of 2 mismatches and up to 3 bp *indels* were allowed in alignment. The assembly rate and quality of individual libraries of 6 samples are detailed in the results section.

Mapped RNA-Seq data were assembled with Cufflink's suite v. 0.17.3 (Trapnell *et al.*, 2010) for normalization and transcript quantification by FPKM values and differential gene expression analysis as explained below.

3.4.3.1a Transcript quantification (counting reads mapped to individual genes or transcripts)

Transcript quantification can be used to estimate gene and transcript expression levels. A number of methods have been developed for the inference of gene and isoform abundance, including RSEM (Li and Dewey 2011), Cufflinks (Trapnell *et al.*, 2010), IsoEM (Nicolae *et al.*, 2011), feature Counts (Liao *et al.*, 2014), and HTSeq (Anders and Huber *et al.*, 2015).

For the present study, Cufflinks was used to process TopHat alignments. After running TopHat (Trapnell *et al.*, 2009), 'BAM' files of reading alignment were submitted to Cufflinks (v. 0.17.3) to assemble transcripts and estimate their abundance for each sample. Then generated a transcriptome assembly for both tropical and temperate carrots root triplicates transcriptome data. Regions of the genome that were covered by sequence reads were treated as potential exons by Cufflinks. Cufflinks used the mapped junctions to assemble transcripts in these covered regions. Cufflinks (v. 0.17.3) generated six files for the individual samples from six TopHat files considered as inputs. These assemblies were then merged using the Cuff-merge utility package. This merged assembly provides a uniform basis for calculating gene and transcript expression in each root transcriptomic data.

The reads and the merged assembly are submitted to Cuff diff, which calculates expression levels and tests the statistical significance of observed changes between the biological samples of the experiment. Cuff-diff learns how read counts vary for each gene across the replicates and uses these variance estimates to calculate the significance of observed changes in expression across the samples under experiment *viz.*, tropical root tissue vs temperate root tissues. Cuff diff also performs an additional layer of differential analysis by grouping transcripts into biologically meaningful groups (such as transcripts that share the same transcription start site (TSS), Cuff diff identifies genes that are differentially regulated at the transcriptional or post-transcriptional level.

Additionally, the abundance of assembled transcripts was calculated and reported as Fragments Per Kilobase of exon per Million fragments mapped (FPKM). This metric is used to compare a single gene's transcript abundance across treatments by normalizing each treatment library's sequencing depth. Confidence intervals were also reported for each FPKM. Cufflinks do not use prior annotation models as it performs the assembly and quantification steps. As a result, novel discoveries including new splice sites, transcriptional start sites, and functional regions of the genome can be identified in this analysis.

Tests for statistical significance (P -values) accounted for multiple testing by computing the False Discovery Rate (FDR), using the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). The FDR adjusts P -values for the number of hypothesis tests performed to control type I error rate (proportion of false positives). Differences in expression were statistically significant when the q -value (FDR adjusted P -value) was less than 0.05. The differentially expressed genes were filtered to identify significant DEGs. The criteria considered for filtering of significant DEG includes $\text{Log}_2\text{FC} \geq 1$ as up regulated (in tropical in other words, down-regulated in temperate) and ≤ -1 as down-regulated (in tropical or in other words up regulated in temperate) genes with $P\text{-value} \leq 0.05$ and $\text{FDR} \leq 0.01$

3.4.4 *De novo* transcriptome assembly (both vegetative and reproductive tissues separately)

Carrot (*Daucus carota* L. ssp. *sativus*), although has a well-assembled genome or transcriptome is currently available; comparison of root transcriptome between *Daucus carota* ssp. *sativus* var. *atrorubens* Alef. (Asiatic/tropical adapted) and *Daucus carota* ssp. *sativus* var. *sativus* (Western/temperate adapted) has not been done earlier. Therefore, we must rely upon a *de novo* assembler for transcript sequence reconstruction and identification of novel transcripts between these two cultivar types for utilization in crop improvement programs. In the present study Trinity v 2.8.6 (Haas *et al.*, 2013) was used for assembly.

The assembler Trinity depends upon three sequential modules; called Inchworm, Chrysalis, and Butterfly for the whole assembly procedure (Fig3). Initially, all overlapping k -mers are carefully chosen from our RNA-seq reads. Inchworm formerly

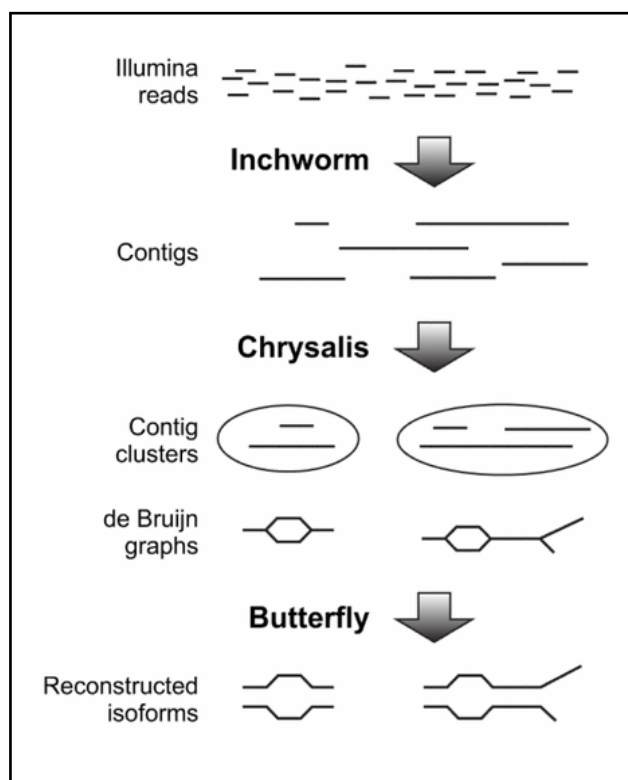


Fig. 3: De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity (Haas *et al.*, 2013).

inspects each unique k-mer in the decreasing order of abundance, and it produces transcript contigs by an extension based on (k-1)-mer overlaps (Grabherr, *et al.*, 2013). Inchworm often creates full-length transcripts for a dominant isoform, but then again it reports just the unique segments of alternatively spliced transcripts; this procedure works well for data that have limited amounts of duplicated sequences, such as transcriptomes.

Chrysalis groups related Inchworm contigs into clusters, which symbolize the full transcriptional complexity for a particular gene or sets of genes sharing sequences in common (Grabherr, *et al.*, 2013). Disjoint de Bruijn graphs are constructed for each of these clusters, followed by partitioning the entire read set from the disjointed graphs.

Lastly, Butterfly processes the disjointed de Bruijn graphs in parallel, reporting full length transcripts for alternatively spliced isoforms and separates transcripts that correspond to paralogous genes (Grabherr, *et al.*, 2013). The butterfly then traces the RNA-seq reads through the graph and determines connectivity based on the read sequence and any related paired-end data (Grabherr, *et al.*, 2013). If those connections were not validated by traced reads, Butterfly then splits the graph into several disconnected sub-graphs and processes each separately. The butterfly then traverses the supported graph paths and reconstructs the transcripts that match the original cDNA data.

Trinity correspondingly supports tools such as Bowtie, RSEM, edgeR and DESeq that take our assembled transcripts and test for differential expression while accounting for both technical and biological sources of variation and correcting for multiple hypothesis testing. After the construction of *de novo* assembly considering all the six cleaned paired-end libraries, transcripts/contigs generated were further used for downstream analysis as discussed below.

3.4.4a Abundance estimation using RSEM (RNA-Seq by Expectation-Maximization -RSEM)

To implement our differential expression analysis in the final stages of our work, we preferred to implement transcript quantification, for estimating the gene and isoform abundances (expression levels) based on the mapping of our RNA-seq reads to the

Trinity assembly. The most obvious complication in evaluating abundances is because RNA-seq reads occasionally do not map uniquely to one gene or isoform, this is where a statistical model is critical for realizing the most accurate estimates.

RSEM uses Bowtie to align the RNA-seq reads of the individual libraries to our Trinity assembly. The essential feature to consider is that the aligner reports all valid alignments of a read and not just a sole best alignment. RSEM utilizes Bowtie to discover all alignments of a read with maximum two mismatches in its first 25 bases by default (Li and Dewey, 2011). This allows RSEM to decide which alignments are most likely to be correct.

After the alignment of reads, RSEM calculates abundance estimations by the Expectation-Maximization algorithm (Li and Dewey, 2011). Expectation-Maximization algorithm provides estimates for reading counts (reads derived from a specific gene), which equates to a gene's expression level. RSEM version 1.2.7 was used for abundance estimation.

3.4.4b Assembly normalization and quality assessment of assembly

The quality of the assemblies was assessed using several methods. First, the assemblies were evaluated against a database of single-copy orthologous genes for plants as implemented in BUSCO v.1.161 (Simão *et al.*, 2015). After BUSCO assessment, the processed RNA-seq reads were mapped to the own combined transcriptomic assembly (*De novo* generated assembly) to determine the percentage of mapped reads, properly paired reads, and singletons. The information obtained in both analyses was used to assess the quality of *de novo* assembly.

Objective 2. Differential gene expression analysis and identification of candidate genes for adaptability and flowering related genes

3.5 Identification DEGs

3.5.1a Secondary storage root transcriptome: Reference-based DEGs

The main goal of functional analysis of DEG is to provide biological insights in the context of the experimental hypothesis that was raised in the beginning. After identification of DEGs, it is important to provide the biological identity to these genes

and gene products through the process of functional annotation. The latest step in a standard transcriptomics study is generally the characterization of the molecular functions or pathways in which differentially expressed genes are involved. Further, the detailed representation of these significant genes for their functions in biological processes, molecular functions, and molecular processes by Gene Ontology (GO) analysis is necessary to describe the body of knowledge within a given domain. DAVID is the most efficient tool for functional annotation and GO analysis among many of the bioinformatics tools available, as it contains annotation data for most model species.

3.5.1b Identification of DEGs (*De novo* -based identification on root and FPL tissues at vegetative and reproductive phases)

The subsequent two sections provide related information regarding two different differential expression analysis tools used for our work. Counts from an RNA-seq experiment are non-negative integers and follow a discrete distribution, so the methods were developed for differential expression analysis using the Poisson distribution and the Negative Binomial (NB) distribution models. The Pair-wise Differential Expression Analysis tool was utilized to perform differential expression analysis of count data arising from an RNA-seq experiment. The calculation of the significance of the change in expression of each of the genes between two samples estimated by both edgeR and DESeq uses a variation of the Fisher exact test adopted for NB (Negative Binomial) distribution and they return exact P values computed from the derived probabilities.

a) edgeR

Our first tool was used for the analysis of differential expression rather than the quantification of expression levels. It is concerned with relative changes in expression levels between both tropical-type and temperate-type root transcriptome sets, but not directly with estimating absolute expression levels.

edgeR adjusts any differential expression analysis for varying sequencing depths as represented by differing library sizes. This is part of the basic modeling procedure and flows automatically into fold-change or p-value calculations. It is always present and doesn't require any user intervention. edgeR can work with expected counts as output by RSEM.

A table of integer read counts, which was produced by Bowtie and RSEM, provides the number of reads that were aligned to each gene for every sample. Sequencing depth or the library size plays an important role in affecting the calculation of reading counts for each RNA sample. More reads are captured from highly expressed genes, and fewer reads will be captured from genes expressed at a lower level. A greater sequencing depth will be required to capture the low abundance transcripts, but at the expense of generating higher counts; therefore, normalization must occur to remove any biases.

To adjust differential expression analysis, edgeR uses differing library sizes for varying sequence depths due to the biological variation between samples (Robinson and Oshlack, 2010). The highly expressed genes can use up a great amount of the total library size (mapped reads), which can cause the remaining genes to become under-sampled (Robinson *et al.*, 2010). edgeR normalized for RNA composition by finding a set of scaling factors for the library sizes that reduced the log-fold changes (quantitative difference between expression levels) between the samples for most genes, which removed biases to achieve a more accurate analysis (Robinson *et al.*, 2010).

Differential gene expression analysis was performed in Trinity (Version) using the package EdgeR (Version 3.11) with FDR correction ≤ 0.05 , P-value ≤ 0.01 , fold change ≤ 2 or ≥ 2 . EdgeR was also used to normalize the expected counts for relative expression and effective library size using the Trimmed Mean of M-values (TMM) normalization method (Robinson and Oshlack, 2010). It uses a weighted trimmed mean of the log expression ratios between samples.

TMM is a between-sample normalization method in contrast to within-sample normalization methods (RPM, RPKM/FPKM, or TPM). TMM will be a good choice to remove the batch effects while comparing the samples from different tissues or genotypes or in cases where the RNA population would be significantly different among the samples (Robinson and Oshlack, 2010). TMM is implemented in edgeR and performs better for between-samples comparisons. Then differential expression for each gene is assessed using Fisher's exact test, but modified to accept replicates.

DEGs filter in edgeR

Genes with very low counts across all libraries provide little evidence for differential expression. From the biological point of view, a gene must be expressed at some minimal level before it is likely to be translated into a protein or to be biologically important. In addition, the pronounced discreteness of these counts interferes with some of the statistical approximations that are used later in the pipeline. These genes should be filtered out prior to further analysis. As a rule of thumb, genes are dropped if they can't possibly be expressed in all the samples for any of the conditions. Users can set their own definition of genes being expressed. Usually, a gene is required to have a count of 5-10 in a library to be considered expressed in that library. Users should also filter with count-per-million (CPM) rather than filtering on the counts directly, as the latter does not account for differences in library sizes between samples.

The counts per gene were normalized to Count per Million (CPM) by dividing it by the total number of mapped reads per sample and multiplying by 1×10^6 (Robinson *et al.*, 2010). The CPM normalized data were then transformed with $\log_2$ using an offset of 1. Genes with at least 1 count per million (CPM) were selected.

b) DESeq

The second analysis tool DESeq normalization method is proposed by Anders and Huber, 2015 and is similar to TMM. It determines differentially expressed genes using a negative binomial distribution to model transcript counts and determine gene-specific variability, similar to edgeR. DESeq uses a negative binomial distribution due to the fact that count data is distinct, skewed or biased, and therefore not approximated sufficiently from a normal distribution. This method gives us the power to more easily detect differential expression and reduce false discoveries by modelling the mean and variance separately (Anders and Huber, 2015).

DESeq and edgeR differ in two areas: normalization and dispersion estimation. DESeq compares the samples against a virtual library created from our RNA-seq count data in which Relative Log Expression (RLE) is estimated. DESeq normalization method estimates RLE scaling factor for each sample to compare the counts obtained

from different samples with different sequencing depth. The counts for each gene in each sample are divided by the geometric mean of the gene across all samples. The median of the ratios for the genes in a sample is estimated as “scaling size factor” used to adjust the total mapped reads count from each sample. In brief, the counts divided by sample-specific size factors determined by median ratio of gene counts relative to geometric mean per gene in DESeq. While, edgeR performs TMM scaling normalization (Grabherr, *et al.*, 2013; Anders and Huber, 2015), that accounts for sequencing depth, RNA composition, and gene length.

The normalization method in both DESeq and edgeR is based on the hypothesis that most genes are not differentially expressed (DE). DESeq and edgeR do not consider gene length for normalization as it assumes that the gene length would be constant between the samples. DESeq and edgeR greatly simplifies the technical influences that need to be taken into account, because any technical factor that is unrelated to the experimental conditions should cancel out of any differential expression analysis. For example, read counts can generally be expected to be proportional to length as well as to expression for any transcript, but edgeR and DESeq does not generally need to adjust for gene length because gene length has the same relative influence on the read counts for each RNA sample.

Both tools also use the Benjamini-Hochberg approach for multiple hypotheses testing to control type I errors. The Benjamini-Hochberg procedure controls the false discovery rate (FDR) which assumes that if one-hundred genes were discovered to be differentially expressed, we expect five to be false discoveries assuming a FDR of 0.05. The probability of having more than five false positives is exponentially reduced. For our research we used an FDR of 0.05.

Only genes represented with an adjusted P-value (FDR) lower than 0.01 and at least a two-fold change were considered as significantly differentially expressed in the pair-wise comparison of the samples.

3.6 Functional annotation of DEGs across cultivars and Gene Ontology of DEGs

3.6.1a Functional annotation of DEGs and Gene Ontology (GO) for Reference-based root transcriptome across cultivars

DAVID is a bioinformatics resource containing an integrated biological knowledge base and analytical tools that aim to systematically extract enriched biological and functional meaning behind a given list of genes or proteins with q-value (Benjamini) < 0.05 . For functional annotation of identified DEGs between temperate and tropical carrot cultivars, initially, a list of significant DEGs were filtered from cuff-diff output file of genes based on significance, $\log_2\text{Fc} (\geq 1 \text{ or } \leq -1)$, P value (0.001). The converted list of gene IDs was imported in to DAVID where the functional annotation was assigned to $>40\%$ of the genes and the annotated genes and functionally classified gene clusters (based on enrichment score) were downloaded for downstream analysis and understanding of biological meaning of the DEGs. The up ($\log_2\text{FC} \geq 1$) and down regulated ($\log_2\text{FC} \leq -1$) DEGs were separately imported to DAVID for functional annotation and GO analysis. The GO categories such as BP, CC and MF for up and down regulated genes were downloaded and compared with other gene functional annotation and downstream analysis tools as discussed below.

3.6.1b Functional annotation of DEGs and GO for *De novo*-based DEGs root and FPL transcriptome across cultivars

The Blast2GO (Conesa *et al.*, 2005) annotation procedure consists of three main steps: BLAST to find homologous sequences, mapping to collect GO terms associated to BLAST hits, and annotation to assign trustworthy information to query sequences. Once GO terms have been gathered, additional functionalities enable processing and modification of annotation results in future works.

The assembled database with longest unigenes was profiled using Blast2GO software to annotate the sequences, determine gene ontology (GO), and identify associated metabolic pathways in the carrot transcriptome. Blast2GO performed a BLAST similarity search on each protein sequence with a significance level cut off of $E-10^{-6}$. Blast2GO compiles candidate GO (Gene Ontology) terms associated with gene

identifiers (GI) of the hits, along with the accompanying evidence codes (EC). Annotation assignments are made by applying an annotation rule that takes into account sequence similarity and node relatedness among the GO term candidates, as well as experimental evidence. Experimental evidence is weighted more highly than electronic evidence in the assignment of GO terms. Enrichment for specific GO terms in the carrot root transcriptome was tested using a module within BLAST2GO that integrates the protein structure comparison tool GOSSIP (Global Structural Superposition of Proteins), to compute Fisher's Exact Test with a default false discovery rate (FDR) of 0.05.

3.6.2 Pathway enrichment analysis (for reference based DEGs in root tissues)

Gene set enrichment (GSE) analysis transforms information from gene expression profiling into a pathway summary. DE genes are quite often involved in the same biological pathways, and GSE results offer greater biological interpretability over individual gene analysis. GSVA (Hänzelmann *et al.*, 2013) extends the current GSE methods to RNA-seq data, and provides increased power to detect subtle pathway activity changes, and constitutes a starting point to build pathway centric models of system biology. GSE analysis was performed with ShinyGO (Ge *et al.*, 2020) by considering carrot as a reference species with the LOG2FC-1, FDR ≤ 0.25 and P value ≤ 0.05 using pGSEA method separately for up and down regulated genes and compared with GO terms obtained for BP, CC, MP in the previous analysis of gene annotation and GO classification to know the enriched pathways in each of these categories.

3.7 Gene Regulatory Network (GRN) analysis/Co-expression analysis

3.7.1 GRN for root storage development and cambium patterning identified for reference based DEGs in roots

Gene expression is mainly regulated by upstream transcription factors (TFs) and usually exhibits spatiotemporal specificity. Specifically, each cell has a particular combination of active TFs and their downstream target genes, which form intricate gene

regulatory networks (GRNs). Thus, GRN profiling is important to understand the mechanisms of gene expression regulation and cellular heterogeneity. In root crops like carrot, storage root development in the context of vascular cambium is critical in deciding the xylem and phloem patterning and in turn yield. The two cultivated orange cultivars used in the present study have a long evolutionary and domestication history and were contrasting in terms of vascular tissue patterning like self-core colour and width of xylem, phloem and cambium tissues. Hence, to understand the role of vascular cambium driven secondary storage root related genes that are specifically involved in vascular cambium driven storage development identified DEGs were subjected to GRN analysis based on the previous report of meta-analysis of storage root development in root crops (Hoang *et al.*, 2020a).

3.7.2 Gene Regulatory Network (GRN) analysis for flower transition between temperate and tropical carrots

With the meta-analysis of genes involved in flower transition partitioning from *Arabidopsis* and their orthology search, few relevant genes responsible for vegetative reproductive phase switch were selected from FPL tissue transcriptome and GRN was generated for assembled transcriptome of FPL tissue comprising both tropical and temperate transcriptome replicates.

STRING v.11b database was used for GRN analysis after converting the carrot gene ids to *Arabidopsis* gene id for secondary storage root developmental genes and floral transition genes.

3.8 Transcriptional factor (TF) identification for root and FPL tissues from *de novo* assembled unigenes

From the assembly was generated from root and FPL tissues by *de novo*-assembly and the fast agile generated was subjected to filtration and redundancy removal. After the removal of redundancy and retaining only the longest isoforms, transcriptional factors were identified in corresponding tissues separately using plant TF

database (PlantTFDB-4.0) by submitting the identified unigenes of root and FPL tissues.

3.9 Genetic variant analysis of root and FPL tissue transcriptome

3.9.1 Identification of microsatellites

De novo assembly was loaded to the MISA tool (Beier *et al.*, 2017) to identify various perfect, imperfect and compound simple sequence repeats. MISA is an ultrafast software that performs with the Fasta file as input file (of any size) and output summary statistics will be generated in the form of csv, pdf and html formats for the assigned tasks like perfect SSR, imperfect SSR (iSSR), compound SSR (cSSR), Variable number tandem repeats (VNTR) and generation of primer pairs (integrated Primer3 tool) for these respective SSR with the multiple parameter options with settings to modify as per the needs.

The criteria followed to generate perfect SSRs were minimum of 15, 12, 10, 6, 6, and 5 mono, di, tri, tetra, penta and hexa nucleotide repeat motifs and for compound SSRs default setting (maximum distance allowed between two SSRs was 100bp) were used. The tables were downloaded and graphs were separately generated in the MS-excel wherever necessary.

3.9.2 Single nucleotide polymorphism (SNP) and Insertion and deletions (*Indels*) discovery

Bam files were converted to vcf files for genetic variant analysis by samtools package to identify SNPs and *indels* distribution across the root transcriptome. High quality genetic variants were filtered from the vcf files with the criteria of Q (Phred score) > 20. Transitions (purine to purine and pyrimidine to pyrimidine) and transversions (purine to pyrimidine and vice-versa) were detected and *indels* with at least 20 Q value were identified for their further utilization in breeding program.

3.10 Identification of candidate genes for root adaptability and flowering related genes

Root adaptability in terms of root development, their physiological fitness with reference to photosynthesis, carbon and lipid metabolism, innate immune response biotic and abiotic stress tolerance under natural field conditions were considered. The functional annotated DEGs were screened for these functions and the filtered candidate genes separately for temperate and tropical carrot cultivars. The storage root tissue specific DEGs for the above-mentioned factors were considered as root adaptability related candidate genes and few of them were also validated with qPCR. GRN analysis as discussed above for the vascular development related genes across temperate and tropical cultivars from the respective candidate genes were also considered as putative candidate genes involved in root adaptation.

For flowering related genes, the candidate genes that were differentially expressed across temperate and tropical cultivars were selected and enlisted with their respective functions. Further, fourteen flowering related genes were selected and validated by qPCR across temperate and tropical cultivars to explain about the expression pattern of flowering related genes.

Objective 3. Functional validation of selected significantly differentially expressed genes by qPCR

Lab experiments including RNA extraction, cDNA synthesis, end-point PCR including gel-electrophoresis and image analysis for confirmation of target genes were carried out at Plant Molecular Biology Lab in the Department of Biotechnology and Crop Improvement. qPCR experiment was done utilizing Step One Plus Real-Time PCR (Thermo Fisher Scientific) instrument maintained at Dept. of Plant Pathology, College of Horticulture, UHS Bagalkot, with the objective of exploring the tropical and temperate carrot candidate genes for breeding and crop improvement of carrots.

Quantitative real-time reverse transcription polymerase chain reaction (qPCR) permits precise high-throughput RNA quantification over a wide dynamic range at a relatively low cost; this technique has high sensitivity and has been widely used for gene expression analysis. In real-time quantitative PCR (qPCR), PCR product is measured at each cycle. By monitoring reactions during the exponential amplification phase of the reaction, users can determine the initial quantity of the target with great precision.

The number of copies of a particular RNA sequence (cDNA) can be determined by measuring the amount of amplified product at each stage during the qPCR cycle. Quantification of amplified products in the present study is obtained using fluorescent DNA-binding dyes (SYBR Green) and 96 well plates and Step One plus Real-Time PCR (Thermo Fisher Scientific) instrument was used.

In order to confirm the reliability and accuracy of the RNA-seq analysis, RT-qPCR was performed on a set of 10 DEGs selected randomly for the set of DEGs identified for in root transcriptome. In these expressions, FC is the fold change ratio between the temperate root-type and tropical group with RPKM expressions.

3.10.1 RNA isolation for qPCR

All steps of analysis followed the MIQE guidelines (Bustin *et al.*, 2009). RNA was extracted from each tropical and temperate carrot sample of root and FPL for validating root and FPL transcriptome respectively. Each sample included three biological and two technical replicates in extraction procedure. Total RNA was extracted by modified TRIzol (Invitrogen; Thermo Fisher Scientific reagent) method. Quality and quantity of RNA was determined using a NanoDrop 2000c (Thermo Scientific) and gel electrophoresis. Purity should be around 1.80 to 2.00 and the RNA integrity was evaluated by 1.5% agarose gel electrophoresis. Stock RNA was stored in the -80°C deep freezer.

3.10.2 cDNA synthesis

cDNA was synthesized from approximately 1,000 ng total RNA using the Prime Script RT reagent Kit with gDNA Eraser (Takara Bio), following the manufacturer's

protocol. cDNA was diluted to 1:40 ratio and used to determine the amplification efficiency (E) and correlation coefficient (R^2) for conducting the qPCR analysis of genes found in root transcriptome in our study. The details of genes selected for root and FPL RNA-seq for qPCR validation is presented in Table 8a, 8b and 8c.

3.10.3 qPCR Protocol

All designed primers for root transcriptome validation and FPL transcriptome data details are provided in Table 9. Primers were designed using Primer 3.0 (Premier Biosoft International, Palo Alto, CA). The specificity of the amplicons was verified by using a single band of expected size in 1.5% agarose gel following electrophoresis.

For Root qPCR experiment, since there were no optimized housekeeping genes in carrot literature for root tissue, five house-keeping genes (Ef-1 α , Actin, Ubiquitin, Tubulin and 18s) were initially used as reference genes selected from the carrot genome for the root samples. The goal here was to identify the stable expressed housekeeping gene as endogenous control for normalization of expression levels across our root tissues used for qPCR. Highly stable reference gene was further utilized in qPCR experiment to study the expression level of 5 randomly chosen target DEGs of root RNA seq experiment. The results of reference gene identification by geNorm, Normfinder, BestKeeper, and the comparative Delta-Ct method are presented in Fig. 4.

The results showed from the concise results of top stable expressed reference genes found to be actin gene followed by Ef-1 α and the same result was also found in Delta CT method, BestKeeper, NormFinder (Fig. 4). However, geNorm showed Actin and ubiquitin are highly stable followed by Ef-1 α . Finally, “Actin” was used as endogenous control in root qPCR experiment to validate the RNA-seq target genes.

For floral primordia and leaf tissue, ubiquitin was already optimized as a good endogenous control in carrot flowering study (Ou *et al.*, 2017). Hence, ubiquitin was used as housekeeping gene to normalize the expression level of target genes in between the samples.

Table 8a: List of genes used for RNA-seq validation of root tissue transcriptome by quantitative real time PCR

Root tissue Transcriptome validation			
Gene_ID	Annotation	Sequence (5' - 3')	Gene_Function
LOC108226637_L	Serine/threonine-protein phosphatase 7 long form homolog	TTGCTTGCGGTTTCCTTGTT	Y2 fine map region
LOC108226637_R		GGCCACCAAGTTGACAAAGC	
LOC108226635_L	Mechano-sensitive ion channel protein 6-like	AACTGGCGGGTCATTTGAGA	Homeostasis and Signalling Pathway
LOC108226635_R		GCCATTGTTGGTGCCTGAAT	
LOC108208449_L	Ubiquitin-conjugating enzyme E2 20-like	GATGGGAGGTGATCCTGGAA	storage root development
LOC108208449_R		TAATCCGCCGGAAAGGAAAG	
LOC108227972_L	Polyubiquitin	TGGAGGACGGAAGGACATTG	storage root development
LOC108227972_R		GATCTGCATTCCACCCCTCA	
LOC108209169_L	Probable WRKY transcription factor 53	CGATTCCGCGATTTTGTGAGA	storage root development
LOC108209169_R		TTTGGTGGTTGTGGACGATG	
LOC108193533_L	Probable glycosyltransferase At5g03795	GGCTTCAAATCGGGAAGGA	Carbon metabolism
LOC108193533_R		CCATGCATGCTACCAGCAAA	
LOC108211283_L	Cysteine proteinase inhibitor B	GGCGAATATTCCGTCGAACA	seed germination
LOC108211283_R		ATTCCGGCCACTACCTGCTT	
LOC108207021_L	ATPase WRNIP1	AAGAGGCTGCGAAACAGCAG	Innate immune response
LOC108207021_R		CCATTCGGTGGCTTTGAGAG	
LOC108216678_L	DNA repair protein XRCC2 homolog	CTTTTCGCCTTGGCAATGTC	Innate immune response
LOC108216678_R		TGCACACCCTTCCACTCCTT	
LOC108200546_L	Histone deacetylase 6	CAACCCGACGGATCTGAATC	Innate immune response
LOC108200546_R		GCTTGGATGTTGGCACCTGT	

Contd..

Table 8b: List of genes used for RNA-seq validation of FPL tissue transcriptome by quantitative real time PCR

FPL tissue Transcriptome Validation			
Primer Name	Gene Name	Sequence (5'- 3')	Gene Function
CONSTANS_L	CONSTANS	TGCGCGCAATGATGAAGCTC	CONSTANS promotes flowering by initiating transcription of the FT
CONSTANS_R		GCACTGCTGTCTGAAACTGCT	
GI_L	GIGANTIA	ACAACGAGACCGAGGAACCC	Activates the floral integrator FT in a long-day afternoon
GI_R		GCACAGACAGCAGCAAGGAC	
PHYA_L	PHYTOCHROME-A	GCGTTGCTGGAAGAGTTTGC	PHYA enhance the stability of CO
PHYA_R		ACCTCCATCTTCACTTCTCAGC	
AP2_L	APETALA-2	CCGTTGAAGAAGAGTCGCCG	AP2 and AP2-like genes inhibit the onset of flowering by repressing expression of FT and SOC1
AP2_R		TGTGCCGTGTCTGAATCCACC	
FKF1_L	FLAVIN BINDING KELCH REPEAT F-BOX1	GGAACCAGCACCTTGTGGC	Blue light promotes the transcriptional activation of CO with the help FKF and GI then induce floral transition
FKF1_R		AACTGAAGCGTCCACCAGAG	

Contd..

Table 8c: List of genes used for validation of flowering related genes by quantitative real-time PCR

FPL tissue Transcriptome Validation			
Primer Name	Gene Name	Sequence (5' - 3')	Gene Function
PHYB_F	Phytochrome B	TCATCGGGAAGGTGCTGGTC	PHYB acts to delay floral transition by destabilizing the CO protein through ubiquitination
PHYB_R		GCACGAACTTTCCGCTCTGG	
LEAFY_F	LATE ELONGATED HYPOCOTYL (LHY)	GCCGAACGTCGTCGCCTC	SPLs promote flowering by initiating the expression of several other transcription factors, such as LEAFY (LFY)
LEAFY_R		CTTCTTCCGTCGTTGCTGCT	
miRNA172_F	Micro RNA 172	TCTGTGGCTGGCAATAGGGC	miRNA172 can subsequently cause a reduction in the AP2 repressor gene transcripts, thus resulting in a switch from transcriptional repression to activation of FT
miRNA172_R		AAGCCGGGTGACAGATCAGC	
DcCOL2_F	CONSTANS-LIKE 2 (COL2)	CAGGCTCAGCCGAAACAA	Plays a central role in regulating the downstream florigen gene by integrating light and circadian clock signals
DcCOL2_R		GCCTTTAGAGGGTCTTGGA	
DcSOC1_F	SUPPRESSION OF OVEREXPRESSION OF CONSTANS 1 (SOC1)	AAGAATCGCCAGAAGACGG	SOC1 also activates floral meristem identity through LEAFY (LFY) and AGAMOUS-LIKE 24 (AGL24)
DcSOC1_R		CGCTTGAATCTCTTTTCTGGA	

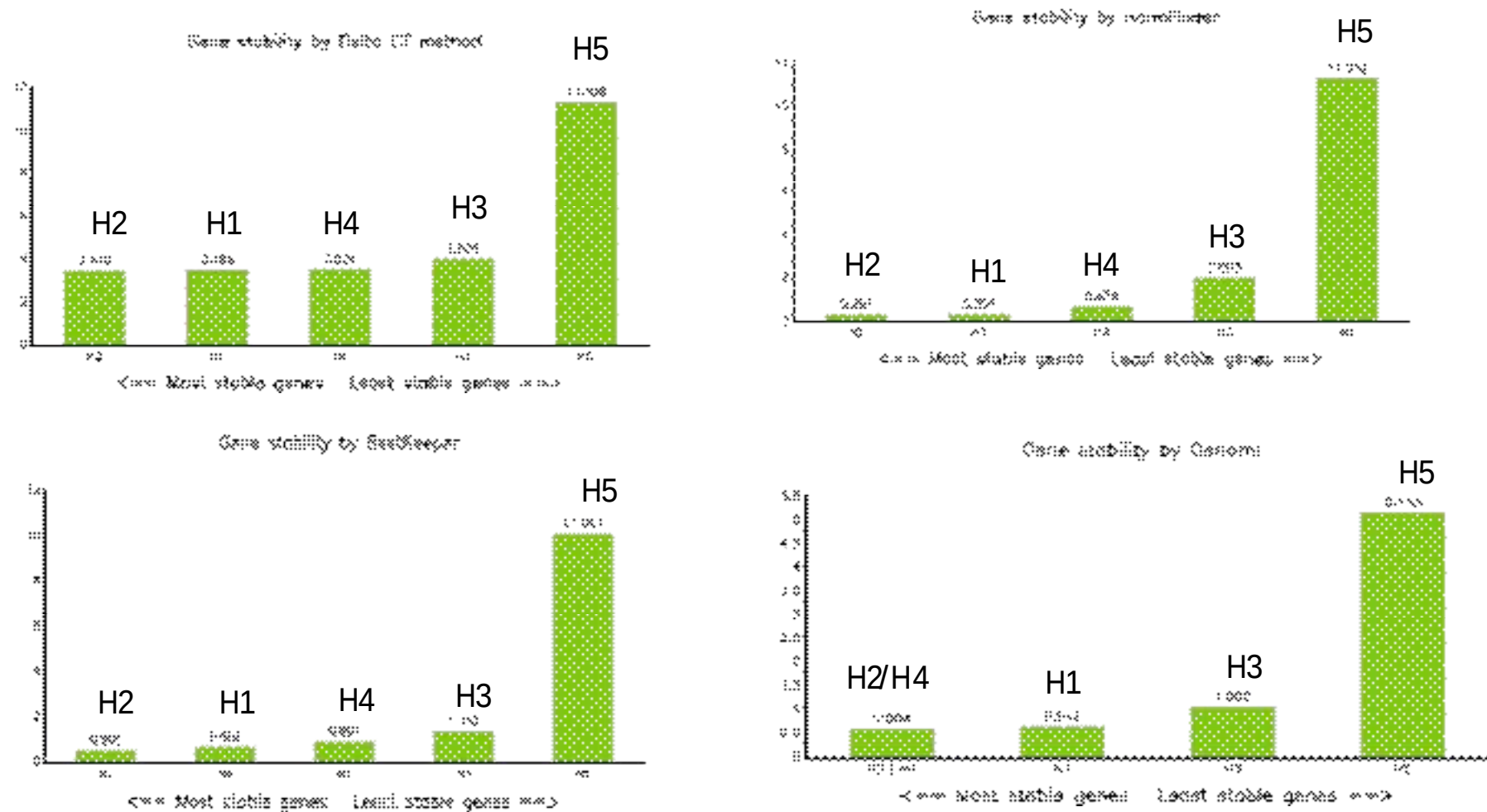
Contd...

Table 8c: List of genes used for validation of flowering related genes by quantitative real-time PCR

FPL tissue Transcriptome Validation			
Primer Name	Gene Name	Sequence (5'- 3')	Gene Function
CCA1_F	CIRCADIAN CLOCK ASSOCIATED 1	GGGCCTTCAAGATGCTTCGC	Circadian Gene- morning phased genes
CCA1_R		CCCTCCAGGCTCCAATTCCC	
TOC1_F	TIMING OF CAB EXPRESSION 1 (TOC1)	CCAGGCGCATGTTAGTCACC	Circadian Gene-evening phased gene
TOC1_R		TTGGCTTATTGCTCGGCTGC	
FT_F	<i>FLOWERING LOCUS T (FT)</i>	ACAGACTGTGTACGCACCCG	FT plays a key role in the floral transition process; promote flowering at the SAM
FT_R		TCTTCGGCCACCTGAACCAG	
FLC_F	FLOWERING LOCUS C (FLC)	CAGACACGGATGATGGCACC	Repress FT expression independently of day length
FLC_R		GCAACTTTAGTGTGCGCTGC	
SVP_F	SHORT VEGETABLE PHASE (SVP)	GCTGAATCGGAGAACGCTGC	Repress FT expression independently of day length
SVP_R		ACCATACGGTAAGCCGAGCC	

Table 9: List of reference genes used for qPCR experiment for validation of root and floral primordial RNA-seq carrot

Gene ID	Gene description	Name of primer	Sequence (5' - 3')	Tm (°C)	PCR efficiency (%)
EF-1α	Elongation factor 1α	EF-1alpha_L	TCAAGGATCTCAAGCGTGGTTATGT	60	100
		EF-1alpha_R	CAGCAATGTGGCAAGTGTGACAAT		
ACT	Actin1 gene	ACTIN_L	CGGTATTGTGTTGGACTCTGGTGAT	60	100
		ACTIN_R	CAGCAAGGTCAAGACGGAGTATGG		
UBQ	Ubiquitin	UBIQUITIN_L	TCTCCGACTCCGTGGTGGTATG	60	98
		UBIQUITIN_R	CTGCCGTCCTCCAACCTGCTTAC		
TUB	Tubulin beta-7 gene	TUBULIN_L	GAGTGGAGTTACCTGCTGCCTTC	60	99
		TUBULIN_R	ATGTAGACGAGGGAACGGAATCAAG		
18S	18S rRNA	18SrRNA	TGACGGAGAATTAGGGTTCTGA	60	98
		18SrRNA	CCGTGTCAGGATTGGGTAATT		



H1: ef-1 α , H2: actin, H3: Ubiquitin, H4: tubulin and H5: 18S

Fig. 4: Selection of most stable endogenous control for qPCR experiment

3.10.4 Quantitative real-time PCR assay

The qPCR was designed according to the minimum information for publication of quantitative real-time PCR experiment, MIQE guidelines (Bustin *et al.*, 2009). Reactions used were PowerUp™ SYBR™Green Master Mix (Applied Biosystem™) in a 10µL reaction volume and were performed in a 96-well plate on Applied Biosystem StepOnePlus™RealTime PCR system (Thermo scientific). Reaction mixtures contained 5µL SYBR Green I Mix, 1µL diluted cDNA solution, 3µL ddH₂O, and forward and reverse primer individually at concentration of 0.5µL each, ROX Reference Dye 0.2 µL for a final 10µL PCR reaction. The following amplification conditions were applied: an initial denaturation step of 95°C for 30 s; 40 cycles at 95°C for 5s; and 60°C for 20s. The final dissociation curve was obtained from 60°C to 95°C to verify primer specificity. Each assay included three technical (for root and 2 technical for flowering genes) and three biological replicates. The general quality assessment of the PCR results was based on the amplification and melting curve profiles of the samples in relation to the assay controls (non-template controls).

3.10.5 Data analysis of qPCR target genes and comparison with RNA seq

The relative expression levels of target genes were calculated using the $2^{-\Delta\Delta Ct}$ method (Quail *et al.*, 2008) with Actin and UBQ as an internal control in storage root and FPL transcriptome. The randomly selected genes used in the qRT-PCR analysis are listed in Table 8a, b and c as discussed above. The RNA pools used in the qRT-PCR analyses were extracted from three independent samples from tropical and temperate carrots storage root and FPL tissue that were different from those used for RNA-seq. All reactions were performed using three biological sample and two technical replicates in both storage root as well as FPL tissue.

3.11 Stability analysis of seven superior temperate/tropical adapted genotypes across seasons and locations

This study was conducted with the objective to identify stable, and adaptable carrot genotypes under various environments of Northern dry zone of Karnataka. Seven

genotypes were tested across three environments during *Kharif* season on randomized block design with three replication (Table 11a). Plot size of 1m × 10 m and 20cm spacing between rows was used. Plants were also evaluated in three seasons at Bagalkot location during 2017, 2018 and 2019. All recommended agronomic practices and managements were applied uniformly. Data were collected on plot basis and analysed with appropriate OPSTAT statistical software for stability parameters.

Seven morphological traits viz., root length (cm), root width (mm), root weight (g), shoot length (cm), number of petioles, shoot weight (g) and shoulder width (g) were recorded from five randomly selected plants for each cultivar per replication. Linear regression model of stability suggested by Eberhart and Russell (1966) was employed and the data was analysed using OPSTAT software.

To analyse the data over three environments and seasons the stability model proposed by Eberhart and Russell (1966) was adopted. The model involves the estimation of three stability parameters like mean, regression coefficient and deviation from regression, which are defined by a mathematical formula (Table 10).

Table 10: Stability parameters for Eberhart and Russell, (1966) model

Regression	Stability	Mean yield	Remarks
b=1	Average	High	Well, adapted to all environments
b=1	Average	Low	Poorly adapted to all environments
b>1	Below average	High	Specifically adapted to favorable environments
b<1	Above average	High	Specifically adapted to unfavorable environments

Regression values of unity are interpreted as average stability

Table 11a. Summary of climatic factors for multi-location trial conducted during the present investigation

Particulars of three environments		Environments		
		Bagalkot (E1)	Arabhavi (E2)	Devihosur (E3)
Geographical position	Latitude (N)	16° 12' N - 16° 46' N	15° 51 '01.30 "N	14°47'59.85"N
	Longitude	74° 59' E 76° 20' E	74° 30'16.81" E	75°23'59.92"E
Soil type		Red sandy loam (Laterite)	Red soils	Red Shallow Soils
Altitude (AMSL)		533 m	836 m	563m
Rainfall average (mm)		585	823	777
Temperature average (°C)		30	31.8	26
Agro-climatic Zone (NARP)		Northern Dry Zone	Northern Dry Zone	Northern Dry zone
Season		<i>Kharif</i>	<i>Kharif</i>	<i>Kharif</i>

Environmental index for each environment across the seasons and phenotypic index for each of the genotype across seasons and locations were estimated to study the suitability of locations/seasons and performance of genotypes respectively with the grand mean for root productivity parameters.

Stability analysis for seven elite lines were performed across the locations (Plate 7, 8 and 9) and seasons following Eberhart and Russel model (1966) was used and for analysis OPSTAT software was used. The information on climatic conditions during the crop season was recorded in Division of Agrometeorology at each experimental site. The details are provided in Table 11a. For fulfilling the objective of stability analysis, multi-season and multi-location trial was conducted for seven elite genotypes of carrot are listed in Table 11b and the list of traits for stability analysis are presented in Table 11c.

Table 11b: List of elite genotypes used for multi-location/multi-season evaluation

S.No	Name of genotype	Tropical/Temperate adapted
1.	UHSBC32-2	Tropical
2.	UHSBC17	Tropical
3.	UHSBC117	Tropical
4.	UHSBC-67	Tropical
5.	UHSBC-23-1	Tropical
6.	UHSBC-34-1	Tropical
7.	UHSBC-145	Temperate

Table 11c: List of traits for stability analysis of carrot cultivars used for multi-location/multi-season evaluation

S.No	Trait
1	Root length (cm)
2	Root width(mm)
3	Root weight (g)
4	Shoot length (cm)
5	Number of petioles
6	Shoot weight (g)
7	Shoulder width (mm)



Experimental site



UHSBC32-2 (G1)



UHSBC17 (G2)



UHSBC117 (G3)



UHSBC67(G4)



UHSBC23-1 (G5)



UHSBC34-1 (G6)



UHSBC100 (G7)

Plate 7: Stability analysis of seven elite carrot cultivars at Bagalkot location (Northern dry zone of Karnataka)



Experimental site



UHSBC32-2 (G1)



UHSBC17 (G2)



UHSBC117 (G3)



UHSBC67(G4)



UHSBC23-1 (G5)



UHSBC34-1 (G6)



UHSBC100 (G7)

Plate 8: Stability analysis of seven elite carrot cultivars at Arabhavi location (Northern Transitional Zone)



Experimental site



UHSBC32-2 (G1)



UHSBC17 (G2)



UHSBC117 (G3)



UHSBC67(G4)



UHSBC23-1 (G5)



UHSBC34-1 (G6)



UHSBC100 (G7)

Plate 9: Stability analysis of seven elite carrot cultivars at Devihosur location (Northern Transitional Zone)

4. EXPERIMENTAL RESULTS

Current knowledge on comparative understanding of genetic makeup underlying storage root adaptability in terms of root developmental genes, defense related candidate genes for biotic and abiotic stress, vascular cambium related genes or photosynthetic efficiency in terms of their source-sink transportation and flowering mechanism of tropical adapted orange/Asiatic type cultivar and temperate adapted orange/Western type of cultivars are scanty. Till now investigations were concentrated on comparing the storage root development of Western type/temperate orange carrots with temperate white colour wild carrot roots (Machaj *et al.*, 2018). Asiatic types/tropical region adapted, orange carrots are least explored in molecular studies and breeding programs. Phylogenetic analysis has clearly indicated that ‘tropical carrot’ has originated from ‘Western Asia’ and ‘temperate carrot’ may be selected from ‘tropical carrot’ (Rong *et al.*, 2014; Iorizzo *et al.*, 2013, 2016; Que *et al.*, 2019).

Research related to reproductive biology of carrot is limited as compared to that of model plants and has remained as inadequate despite the availability of powerful genetic research on carrot and their wild relatives over the last decades. There are no reports on comparative understanding of flowering pathway of both late flowering (Western/temperate type) and early flowering domesticated cultivars (Asiatic/tropical type) of carrot in response to different photoperiodic conditions. Therefore, presently generated transcriptome data can provide greater biological insights w.r.t. both root and reproductive biology of early flowering tropical adapted and late flowering nutritionally rich temperate adapted carrots. Considering these facts, the present investigation was planned to have detailed insight into **“Transcriptome profiling of temperate and tropical adapted carrot (*Daucus carota* L. subsp *sativus*) cultivars to decode the candidate genes related to root adaptability and flowering”**.

Until today no transcriptomic research has been done to investigate the molecular mechanisms underlying temperate and tropical adapted storage root and flowering biology together in carrot by examining roots and leaves together with floral primordia in a single experiment. In view of domesticated temperate orange as a biennial type, tropical orange as an annual type crop, in which root is the edible part and

seed is the propagating material, it is most relevant to study the root adaptability, vascular cambium patterning and flowering transition in these contrasting cultivars.

Root yield of tropical carrots are comparatively higher, but xylem and phloem patterning/differentiation is problematic in tropical cultivar, whereas ‘self-core’ with uniform orange is the useful characteristic feature of temperate cultivars. Root adaptability in carrot is the ability of a plant to produce edible root with higher root yield, uniform root length, tolerance to biotic and abiotic stresses. Ability to induce flowering and successful production of seed set under varied climatic conditions is also another important criterion of adaptability. Flowering behaviour of temperate and tropical adapted carrots is highly contrasting in terms of vernalization and photoperiod requirement, as they are classified into biennials and annuals respectively. Hence, study of two contrasting carrots such as temperate adapted/Western type and tropical adapted/Asiatic types that differ in growth habit, yield, and ability to adapt and produce superior roots and seeds after flowering is highly useful. The two economic tissues such as secondary storage root at vegetative phase and leaf along with floral primordia (FPL) at reproductive phase were considered to be important for RNA-seq in the present work. Transcriptome profiling of these two tissues provide greater information regarding various candidate genes involved in root development, photosynthetic efficiency, carbohydrate metabolism, biosynthetic processes, constitutive defense responsive genes for biotic or abiotic stresses and reproductive transition, flowering related genes.

Throughout the chapter, Asiatic domesticated orange cultivar is mentioned as “tropical type” and Western domesticated orange carrot cultivar is mentioned as “temperate type”.

4.1 Root adaptability/root plasticity parameters of temperate and tropical carrot cultivars

Root phenotypic plasticity in terms of growth and yield components of tropical and temperate adapted carrots was evaluated under warmer climate *i.e.* Bagalkote during 2017-20. Root plasticity parameters like root length, root width, phloem (cortex) width, xylem width, hairiness, lateral branching, plant height, number of petioles and yield parameters, beta carotene, TSS and flowering ability under natural field were recorded as listed in Table 12.

Table 12: Morphological traits distinguishing tropical and temperate carrot cultivars

Traits	Tropical cultivar (UHSBC-23-1)	Temperate Cultivar (UHSBC-100)
	Mean	Mean
Plant height (cm)	68.13±1.75 ^a	53.76 0.73 ^b
Root Length (cm)	16.52± 0.98 ^a	13.18±0.26 ^b
Shoot length (cm)	52.49±1.38 ^a	46.64±3.21 ^b
Shoulder Width (mm)	21.53±1.03 ^a	16.22±0.69 ^b
Root diameter (mm)	26.07±1.61 ^a	25.39±0.52 ^b
Phloem width (cm)	0.50±0.12 ^a	0.85 ±0.14 ^b
Xylem width (cm)	0.82±0.08 ^a	0.40±0.01 ^b
No. of petioles	7.91±0.25 ^a	8.46±0.15 ^b
Shoot weight (gms)	70.91±5.45 ^a	24.49±0.96 ^b
Root weight	72.70±5.88 ^a	40.66±4.11 ^b
Plot yield (100 plants) Kg	6.69±0.62 ^a	4.46±0.34 ^b
R/S ratio (Length)	0.32±0.02	0.31±0.01 ^b
R/S ratio (Weight)	1.39±0.15	1.83±0.12 ^b
TSS ⁰ Brix	7.08±0.11 ^a	6.86±0.14 ^b
β-carotene mg/100mg	8.5±0.16 ^a	6.5±0.12 ^b
Root Colour (External)	Orange	Deep orange
Phloem colour	orange	Deep orange
Xylem/core colour	Yellow	Deep orange
Root texture	Coarse	Smooth
Root shape	Tapering	Cylindrical
Leaf type	Norma	Fern
Root hairiness	Moderate	Absent
Days to maturity	Early (80-85 days)	Late (90-95 DAS)
Flowering ability in tropics	>80 % Flowering	Absent

*Two samples, t-test assuming equal variance was used to test the significant difference between temperate and tropical cultivars for individual trait at Significance 0.05

The highest phloem width was recorded in the temperate cultivar (0.86 mm) with uniform pigmentation and almost complete absence of xylem patterning, whereas, the tropical cultivar showed the highest xylem width of 0.84 mm with least phloem 0.50. Exterior and interior colours are important characteristics receiving attention in all the tropical carrot breeding programs. The average shoot weight (70.91 ± 12 g) of the tropical cultivar single plant was higher than the average shoot weight of 24.44 ± 3.46 g in temperate cultivar. Whereas, in temperate adapted cultivar, average root weight (40.66 ± 4.11) was lower than that of average root weight (72.70 ± 5.88) of tropical adapted Asiatic cultivar (Table 12). Western type had significantly higher carotenoid content with the mean β -carotene content 8.5 mg/100mg in temperate to 6.5 mg/100mg in tropical cultivars; and TSS ranged from 6.86 °Brix. to 7.08 °Brix in temperate and tropical cultivars respectively.

The following sections will provide a detailed description of transcriptome profiling computational analysis results related to reference-based transcriptome assembly and *de novo* transcriptome assembly of root tissue and *de novo* assembly of floral primordia + leaf (**FPL**) tissue. Further, differential expression analysis pipeline results for identification of candidate gene between two cultivars, gene ontology (GO) categorization, gene regulatory network (GRN) construction for root vascular cambium genes and flowering genes identified in the present study is also presented. Further, the selected significant candidate genes were validated by qPCR expression study. Later few superior genotypes identified in the previous research were subjected to multi-location and multi-season evaluation in the present study to identify the stable genotypes by stability analysis. The details of the results are presented below.

4.2 Generation of raw reads by transcriptome profiling of roots and FPL tissues in temperate and tropical cultivars

a) Pre-processing and Quality filtration

Raw paired-end reads (2×100 bp for roots and 2×151 bp for flowering tissues) were obtained from two cultivars with three biological replicates each phase comprising of six paired-end libraries. A total of twelve RNA-seq paired end libraries were sequenced in the present investigation from two cultivars and two tissues in three

replicates. The reason for selection of pooled tissues of floral primordia and leaf tissues are mentioned in Material and Methods.

The details of the experimental results are discussed in the following subheadings. Since there are two tissues taken at two developmental stages and analysis were done independently, hence, the results are discussed for root transcriptome and FPL tissues independently one after other.

b) Filtering and trimming of raw reads and generation of cleaned reads of Root and FPL tissues

Three biological replicates from each cultivar were processed to construct six root samples paired-end RNA libraries. A total of 140 million (M) paired-end reads were generated from root transcriptome comprised of six libraries, ranging from 19.09 to 29.10 M high quality paired-end raw reads in individual libraries (Material methods section). The quality of the raw reads was examined using FastQC v0.10.0 and analysed data included various parameters such as good quality basic statistics, per base sequence quality, sequence length distribution and GC content. Adapter content, sequence duplication level, low quality reads, erroneous base calls, or overrepresented sequenced were trimmed and filtered for each sample individually by using default parameters Trimmomatic v0.36 (Bolger *et al.*, 2014), which discarded data with poor quality scores ($Q < 25$), retained read length minimum of 36 bp for roots and 100bp for FPL tissue. A total of 2 seed mismatches were allowed in trimming in both transcriptome data.

4.3 Cleaned raw reads root and FPL transcriptome used for assembly

a) Cleaned reads for Assembly of root transcriptome

Only the cleaned reads generated from Trimmomatic were utilized for downstream analysis. Table 13 indicated the survived and dropped reads from the Trimmomatic generated output file. A total of 130.62 million (M) (92.71%) high quality (Phred score >25) paired end reads were retained; whereas, 10.26 M reads (7.29 %) were discarded.

having poor quality reads were removed from the downstream analysis. From each library, 17.63 M to 24.93 M reads per sample were used in both reference based and *de novo*-based assembly of root transcriptome data. A summary of the raw data generated, and the trimmed for root transcriptome, is summarized in Table 13.

b) Cleaned input reads for Assembly of FPL (Floral Primordia pooled with Leaf) transcriptome

Three biological replicates from tropical and temperate carrot types were processed to construct six FPL libraries. Adapter sequences used during the library preparation and low-quality raw reads were trimmed from the Novaseq reads using a Java-based program Trimmomatic v0.36 (Bolger *et al.*, 2014). Adapters and low sequence reads were removed from raw reads. The adapter-free reads were further filtered based on the quality. The parameters for quality trimming were set for a minimum mean quality of Q25 across the read and to trim low-quality bases at 3' end. The minimum read length of read lengths 50 bp for the Novaseq reads was used as parameters trimming and filtering. A total of 202.36 million (M) (86.74%) high quality (Phred score >30) paired end reads were retained; whereas, 30.93 M reads (13.26 %) having poor quality reads were removed from the downstream analysis. From each FPL library, 20.24 M to 41.94 M reads per sample were used in *de novo*-based assembly of FPL transcriptome data. A summary of the raw data generated, and the trimmed reads for FPL transcriptome, is summarized in Table 13.

4.4 Assembly/alignment of root (Reference based and *de novo*-based) and FPL tissue (only *de novo*-based) transcriptome

Alignment or assembly of sequenced reads to reference genome is a core step in RNA-seq analysis for further downstream analysis such as transcript quantification and DGE analysis. In the present study, both reference-based and *de novo* assembly was performed for root transcriptome data as high-quality reference carrot genome was available for the former analysis and to identify cultivar and tissue specific transcripts identification for the later analysis.

Table 13. Summary of input reads, filtered and trimmed reads of six samples of two cultivars of carrot root and FPL tissue transcriptome data

Samples	Root transcriptome (vegetative phase)-100bp PE				
	Input Reads	Surviving Reads	% Survived reads	Dropped Reads	% Dropped Reads
Temperate_R1	2,67,96,143	2,49,34,401	93.05%	24,940	0.09%
Temperate_R2	1,96,98,951	1,81,42,740	92.10%	18,653	0.09%
Temperate_R3	2,26,04,096	2,12,26,283	93.90%	27,171	0.12%
Tropical_R1	2,90,96,070	2,67,66,083	91.99%	29,869	0.10%
Tropical_R2	1,90,89,333	1,76,30,817	92.36%	18,729	0.10%
Tropical_R3	2,35,98,359	2,19,18,109	92.88%	23,736	0.10%
Total reads	14,08,82,952	13,06,18,433		1,43,098	
Samples	FPL tissue transcriptome (Reproductive phase)- 151bp PE				
	Input Reads	Surviving Reads	% Survived reads	Dropped Reads	% Dropped Reads
Temperate_R1	3,17,26,236	2,62,23,523	82.66%	55,02,713	2.36%
Temperate_R2	4,82,78,351	4,19,39,945	86.87%	63,38,406	2.72%
Temperate_R3	4,46,34,593	3,89,74,029	87.32%	56,60,564	2.43%
Tropical_R1	3,58,56,048	3,16,59,119	88.30%	41,96,929	1.80%
Tropical_R2	3,48,61,066	3,09,48,384	88.78%	39,12,682	1.68%
Tropical_R3	3,79,36,428	3,26,14,387	85.97%	53,22,041	2.28%
Total Reads	23,32,92,722	20,23,59,387		3,09,33,335	

Note - Total no of trimmed & cleaned reads (Root Tissue) = $130618433/140882952 * 100 = 92.71\%$ data (high Quality data)

Total no of trimmed & cleaned reads (FPL-Tissue) = $202359387/233292722 * 100 = 86.74\%$ data (high Quality data)

Total no of removed reads (Roots tissue) = $10264519/14082952 * 100 = 7.29\%$ data (removed low Quality reads)

Total no of removed reads (FPL) = $30933335/233292722 * 100 = 13.26\%$ data (removed low Quality reads)

*PE- Paired End

4.4.1 Reference-based (TopHat) assembly for root transcriptome

As Bowtie does not allow alignments between a read and the genome that contain large gaps; it cannot align reads that span introns. Hence, TopHat was used in the present work to address this constraint. Bowtie cannot align smaller pieces called segments. These pieces are processed independently for aligning to the genome. When several of a read's segments align to the genome far apart (e.g., between 100 bp and several hundred kilobases) from one another, TopHat infers that the read spans a splice junction and estimates where that junction's splice sites are located. By processing each 'initially unmappable' read, TopHat can build up an index of splice sites in the transcriptome without a prior gene or splice site annotations.

4.4.1.1a Cleaned reads of root transcriptome

In the reference-based assembly, after filtering duplicate and low-quality reads, 130618433 (92.71%) of high-quality cleaned reads were retained for downstream analysis. The **high-quality assembly sequence** of *Daucus carota* L. subsp *sativus* (ASMcarrotsv1) was downloaded from NCBI and index was generated from Bowtie. TopHat v 2.0.6 (Trapnell *et al.*, 2009) was used to map trimmed clean reads to the *Daucus carota* reference genome retrieved from NCBI.

A summary of the cleaned reads mapped to a reference carrot genome is presented in Table 14. The mapping rate ranged from 71.6 to 82.8% in temperate (Western type) and 68.5 to 72.7 % in tropical (Asiatic type) carrot across three biological replicates in the respective cultivars. An average mapping rate of 73.87 % was observed for the root transcriptome. This showed that the mapping quality was good enough to conduct the subsequent analysis.

4.4.1.1b Transcriptome reconstruction

It is necessary to pool the data and assemble it into a comprehensive set of transcripts before proceeding to differential analysis. A transcriptome reconstruction for each replicate and the merged transcriptome assembly was produced by Cufflinks and cuff merge packages respectively from the Cufflinks suite v1.3.0 (Trapnell *et al.*, 2012).

Table 14. Summary of the cleaned reads of root transcriptome aligned to the reference genome by TopHat v2.0.6

Sample name	Left mapped reads rate			Right mapped reads rate			TopHat Mapping rate (%)
	Input Read	Total mapped reads	% Mapped reads	Input Read	Total mapped reads	% Mapped reads	
Temperate_R1	24934401	21882512	87.8	24934401	16094457	64.5	76.2
Temperate_R2	18142740	13954891	76.9	18142740	12009046	66.2	71.6
Temperate_R3	21226283	17835932	84.0	21226283	17311001	81.6	82.8
Tropical_R1	26766083	22581408	84.4	26766083	16356661	61.1	72.7
Tropical_R2	17630817	14595612	82.8	17630817	10592745	60.1	71.4
Tropical_R3	21918109	17506456	79.9	21918109	12508173	57.1	68.5

Cufflinks generated corresponding *gtf* files for individual libraries from the corresponding *'bam* files. Six *gtf* files were merged into one *gtf* in the cuff-merge tool.

The resultant cuff merge output file along with the individual sample's assembly files were subjected to differential genes identification between the two cultivars under question *i.e.*, temperate adapted (Western cultivar) and tropical adapted (Asiatic cultivar) carrot roots as explained below.

4.4.1.1c Expression Normalization

Cufflinks comprises a separate package, Cuff-diff, which subtracts the bias in RNA-seq read distribution across each transcript environments to provide abundance estimate through FPKM estimation; fragments per kilo base of transcript per million mapped reads. Cuff diff determines expression in two different samples and tests the statistical significance of each observed change in expression between them. Cuff-diff understands in what way read counts vary for each gene across the replicates and uses these variance estimates to calculate the significance of observed changes in expression. Cuff-diff normalizes expression estimates of each sample by the total number of mapped fragments in the sample then adjusts expression assessments by gene or transcript length, which in turn enables both inter- and intra-sample comparisons.

4.5 Identification of DEGs for root transcriptome across cultivars in reference-based assembly

Comparison of expression levels of genes or other genomic features between different biological conditions/samples is referred to as DGE analysis. DGE is crucially based on the principle of comparison between samples through univariate assessments like t-test in cuff-diff or Fisher's exact test in DESeq and EdgeR. In this study, we use cuff-diff for DGE identification between temperate and tropical roots.

A total number of DEGs and isoforms identified were filtered with $p\text{-value} \leq 0.01$, $\log_2\text{FC} \geq +1$, and $\log_2\text{FC} \leq -1$ for root transcriptome. We identified 3544 DEGs from a total of 35813 genes identified in cuff-diff. Further, these 3544 DEGs were filtered for up and downregulated genes based on $\log_2\text{FC}$ and P value < 0.01 . The genes were declared as upregulated in tropical or in other words downregulated in temperate

root when Log_2FC was ≥ 1.0 and down regulated in tropical in other words upregulated in temperate roots for genes with $\text{Log}_2\text{FC} \leq -1.0$. From the filtering, a total of 1818 and 1726 up and down regulated genes respectively were identified in pair-wise differential analysis as presented in Fig. 5. The filtered list of significant genes, and isoform differentially expressed between two carrot cultivars roots are presented in Appendix I along with the corresponding FPKM values.

Among 3544 DEG, 1818 genes were up-regulated in tropical carrot storage root tissue than temperate, and 1726 genes were found to be down-regulated in tropical adapted storage root (Table 15). In other words, 1818 were down regulated in temperate and 1726 were upregulated in temperate carrot storage roots of carrot Table15. When, a more stringent $\log_2\text{FC}$ value between ≥ 2 to ≤ -2 , a total of 1186 DEGs were significant with P value ≤ 0.01 . In this analysis, 592 DEGs were identified as up-regulated in tropical or down regulated in temperate and 594 DEGs were down-regulated in tropical or up-regulated temperate adapted carrot roots. Fig. 6a represents the volcano plot, and significant genes satisfying $p\text{-value} \leq 0.05$ $\log_2\text{FC} \geq 1$, and $\log_2\text{FC} \leq -1$ are labelled and shown in red and blue dots for up and down regulated genes respectively. The gene query can be also presented as a list of chromosomal coordinates. For each chromosomal region we extract all genes that are at least partially located in the given region. Analysis of genes in chromosomal regions is a useful feature for analysing DEGs distributed across genome as presented in Fig. 6b.

Table 15: List of DEGs identified and annotated for reference-based root transcriptome

Direction of Gene expression	Differentially Expressed Genes	Annotations
Upregulated	1818	799 (43.95 %)
Downregulated	1726	742 (42.99 %)
Total	3544	1541 (43.49 %)

4.6 Functional annotation of DEGs and gene functional classification

Preliminary handling of the data involved modifying specific gene symbols from probe IDs with the assistance of an Entrez Gene ID converter (Alibés *et al.*, 2007) in DAVID tools. For the functional annotation of DEGs, DAVID v6.8 web server was used to gain biological insight into our gene lists, identify enriched gene ontology terms, and found functionally related gene groups.

From the 3544 DEGs of root transcriptome subjected to annotations, 1541 (43.49%) were functionally annotated in David (Appendix-I). Among tropical upregulated genes (1818), 799 genes (43.95%) were annotated and 20 gene clusters or functional gene sets (FGS) were obtained based on gene functional classification with the enrichment score ranged from 0.03 lowest score to the highest score of 3.04 (Table 16). Whereas, 742 (42.99%) out of 1726 down regulated genes were successfully annotated in DAVID with 13 gene clusters/FGS with the enrichment score ranging from 0.06 to 5.87 (Table 16). Remaining 2,003 genes were either uncharacterized or not assigned any function. The details of each FGS are presented in Appendix-I.

Among up regulated FGS, the major group consisted of 222 genes in Group 2 with the related terms associated with the ‘innate and adaptive immune response’. Whereas group1 consisted mainly of ‘probable WRKY transcription factors’ (enrichment score 3.04), Group 14 with six disease resistance response genes (Enrichment score 0.568). Group17 was another major group consisted of 26 genes with ethylene responsive TF and other regulatory genes (enrichment score 0.435).

Whereas, the gene functional classification of down regulated genes in tropical cultivar or up regulated genes in temperate roots were classified into 13 functional gene set (FGS) with the highest enrichment score of 5.866 in Group 1 consisting of 8 genes in which majority were ‘oxygen evolving enhancer protein’. In Group 2 majorities were showing chlorophyll a-b binding protein. Highest number of genes was clustered with the Group 11 (enrichment score of 0.156) and majority of them were related to ‘transmembrane helix’, ‘carotenoid integral component of membrane gene-terms’. The Group 4 consisted of 44 genes with the enrichment score of 1.171 with majority of gene terms related to ‘nucleus’, ‘DNA binding transcription’ and ‘transcription regulation’.

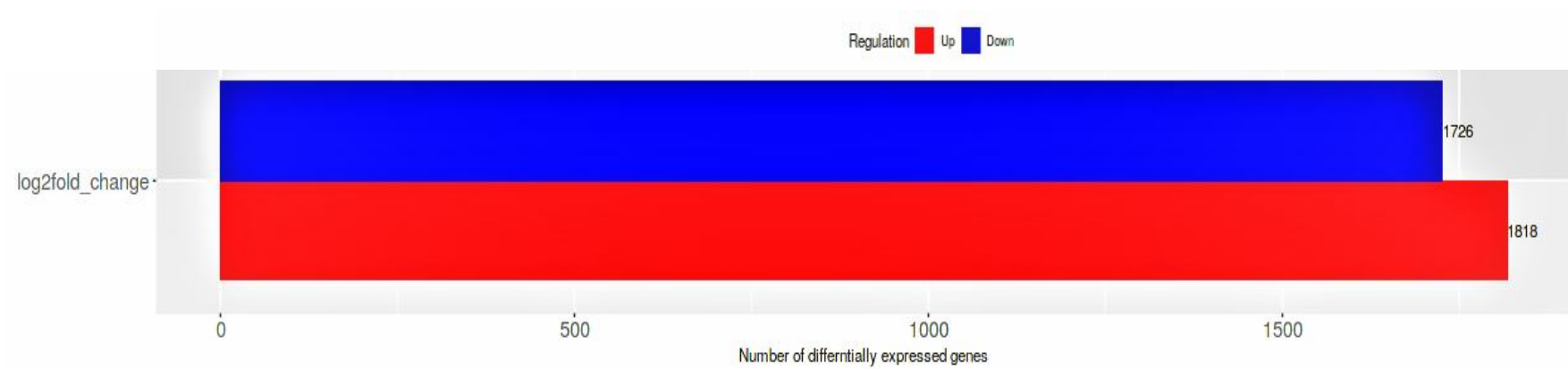


Fig. 5: Number of up and down regulated DEGs identified in pair-wise differential analysis in Cuff-diff package of Cufflinks suite v1.3.0

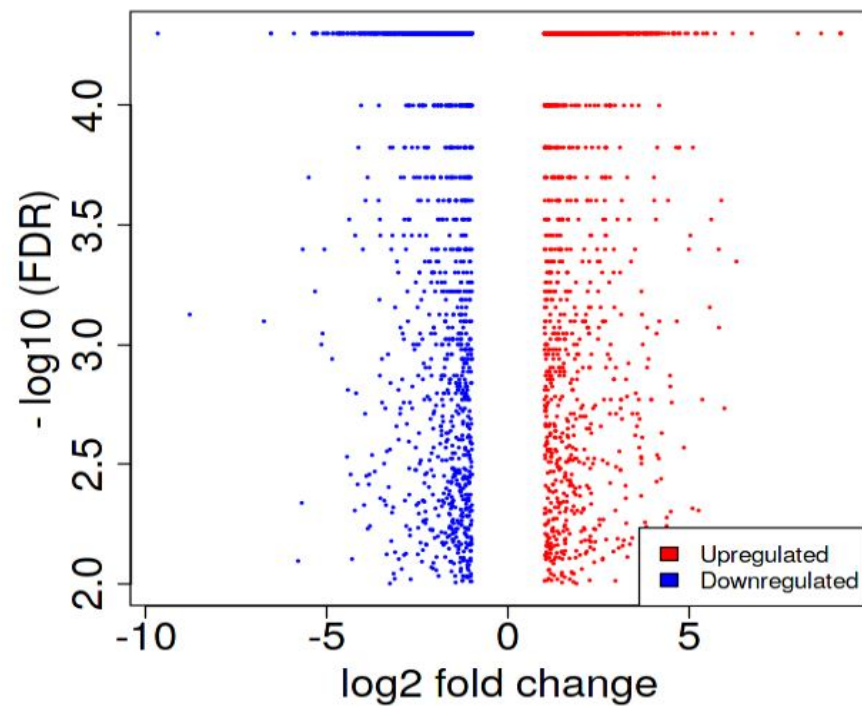


Fig. 6a: Visualization of total DEGs identified by reference-based root transcriptome between tropical and temperate carrot cultivars by volcano plot

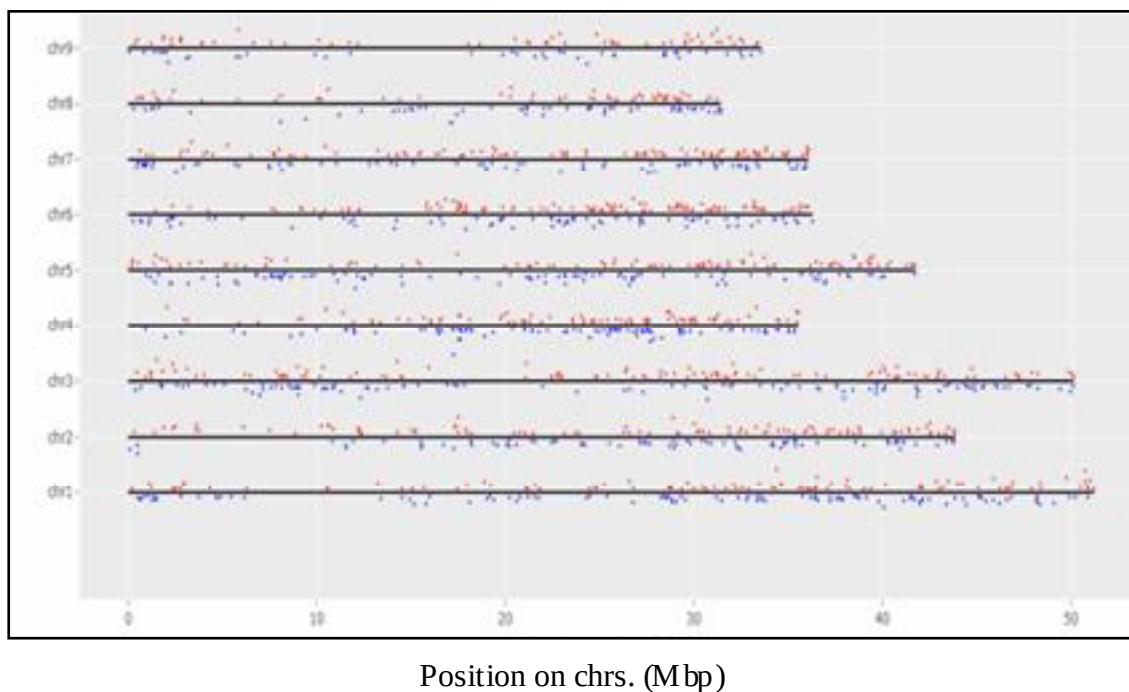


Fig. 6b: Chromosome-wise distribution of DEGs identified for reference-based transcriptome between temperate and tropical root tissues based on FDR (<0.50) and Log2FC (<-1 to >1). Red dot indicates upregulated genes and blue dots indicates downregulated genes

Table 16: Gene functional classification of Up and Down regulated genes with corresponding enrichment score and gene counts extracted from DAVID v 6.8

UP Regulated Gene Clusters	Enrichment Score	Gene Counts	Down Regulated Gene Clusters	Enrichment Score	Gene Counts
Gene Group 1	3.0358	13	Gene Group 1	5.866255	8
Gene Group 2	2.4273	222	Gene Group 2	4.208765	10
Gene Group 3	2.3291	6	Gene Group 3	1.321961	6
Gene Group 4	2.2058	10	Gene Group 4	1.171067	44
Gene Group 5	2.1841	5	Gene Group 5	1.125687	7
Gene Group 6	1.9624	10	Gene Group 6	0.681763	4
Gene Group 7	1.4033	6	Gene Group 7	0.590238	4
Gene Group 8	1.1612	5	Gene Group 8	0.585988	4
Gene Group 9	1.0527	4	Gene Group 9	0.234972	5
Gene Group 10	1.0474	4	Gene Group 10	0.173066	4
Gene Group 11	1.0455	11	Gene Group 11	0.15646	152
Gene Group 12	0.8679	5	Gene Group 12	0.086252	4
Gene Group 13	0.7928	6	Gene Group 13	0.057736	4
Gene Group 14	0.6997	6			
Gene Group 15	0.6403	4			
Gene Group 16	0.5681	6			
Gene Group 17	0.4354	26			
Gene Group 18	0.4316	6			
Gene Group 19	0.1407	4			
Gene Group 20	0.0341	6			

4.7 Gene ontology (GO) categorization

GO provides a hierarchically organized set of thousands of standardized terms for biological processes, molecular functions and cellular components, as well as curated and predicted gene annotations based on these terms for multiple species. Biological process GO annotations are the most commonly used resource for pathway enrichment analysis.

Up and down regulated genes identified from reference-based assembly were subjected to gene ontology classification by categorizing the functional annotated genes into biological process (BP), cellular component (CC), molecular functions (MF) and KEGG pathway classifications. A popular database, iDEP (Ge *et al.*, 2018), an integrated web browser for differential expression and pathway analysis of RNA seq was used for Gene ontology categorization, phylogeny construction of GO categories, Gene Set Enrichment Analysis (GSEA) by submitting the total DEGs (3544) with the corresponding P-value (<0.01) and Log2FC (>1 or <-1) as input in csv format to achieve a broader functional characterization.

As a result, 3544 DEGs in the root transcriptome were classified into 45 functional subcategories within three main categories (BP, CC and MF) (Fig. 7a-b, 8a-b, 9a-b). In total, 1018 DEGs in the roots were associated with BP terms, 392 root DEGs were associated with MF terms, 519 root DEGs annotated with CC terms and 427 root DEGs were associated with KEGG pathway. It is important to note that in many cases the same gene can be assigned to more than one category.

In GO terms, the most represented BP subcategories in tropical down regulated were “Regulation of metabolic process” (82 gene counts), followed by “Regulation of macromolecular metabolic process (77 genes) and cellular metabolic process (76) aromatic compound biosynthetic process” and “Photosynthesis” (43). Whereas, in tropical upregulated genes “localization (94 genes) and transport (93 genes)’ were major BP identified (Fig. 7a). Fig. 7b showing Enrichment tree relationship of Gene Ontology terms identified for Biological Processes (BP) set of genes up-regulated for Localisation

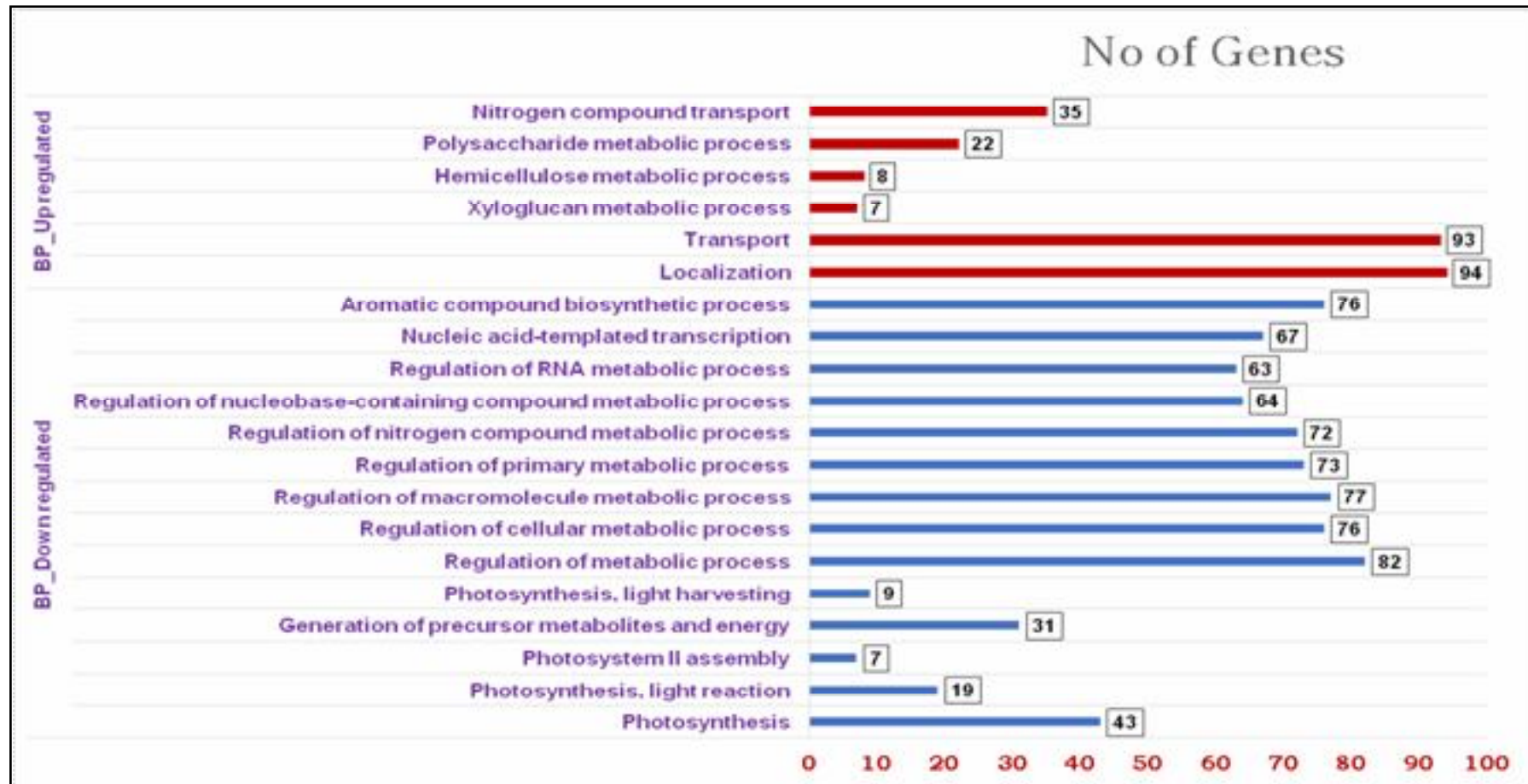


Fig. 7a: Gene Ontology terms identified for Biological Processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93). Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value



Fig. 7b: Enrichment tree showing relationship of Gene Ontology terms identified for Biological Processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root. A Value indicates the P value associated with corresponding GO terms (iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and up regulated, brighter green is highly down regulated) and the size of the terms represents the number of genes that are significant from our list.

and transport were clustered together with the corresponding p value of 1×10^{-4} along with Nitrogen compound Transport with p-value of 9×10^{-3} , then genes involved in Hemicellulose (p-value- 2×10^{-3}), Xyloglucan, and Polysaccharide metabolic process. Contrastingly, set of genes responsible for photosynthesis, light reaction and assembly complex were down-regulated and grouped together.

Among the CC terms, for upregulated tropical cultivar GO terms included “integral component of membrane (196) and two other terms with few gene counts such as apoplast and cell wall 6 and 12 genes respectively. Among the downregulated DEGs of tropical cultivar for CC terms, “Plastid” (74) was the most dominant sub-category, followed by chloroplast (71), “Thylakoid” (55) and ten others (Fig. 8a). Fig. 8b exhibit gene enrichment tree showing relationship of GO terms identified for Cellular components (CC) are photosystems I and II, Photosynthetic membrane, Chloroplast (71), Plastid (74) and Thylakoid (55), Photosynthetic membrane (45), Thylakoid membrane (39) etc. were found to be down-regulated in tropical adapted carrot though genes involved in integral component of membrane (196), cell wall (12), Apoplast (6) were up-regulated. Up-regulated genes were not enriched for any of the cellular component GO terms in iDEP, but DAVID identified few terms.

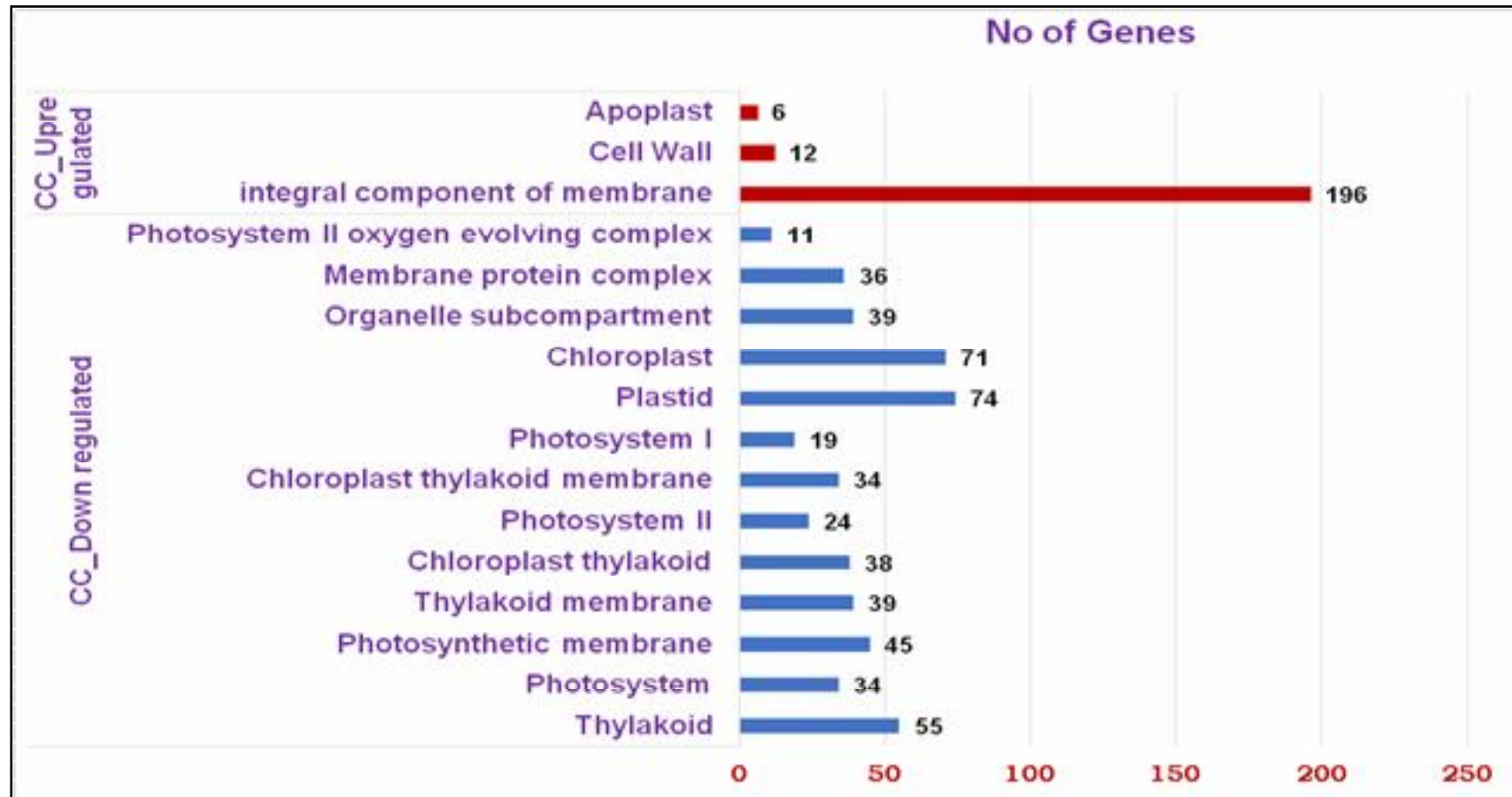
As for MF, the major upregulated GO terms for tropical cultivar were molecular function regulator (52), transcription regulator (39), DNA binding transcription activity (38) and three others. Whereas, among major downregulated MF categories in tropical cultivar were “DNA binding (93 genes)”, “Molecular function regulator” (49 genes) and four others (Fig. 9a). The enrichment tree shows the related GO terms with reference to MF. Additionally, Fig. 9b illustrate GSEA tree for noticing relationship of GO terms identified for Molecular Functions (MF) associated with up-regulated genes mainly involved in transcription regulator activity, DNA binding transcription factor activity, molecular function regulator, sequence specific DNA binding were gathered together and genes involved in Xyloglucan transferase activity and glycosyl transferring groups were grouped together. While, downregulated GO terms of tropical cultivar were DNA binding, chlorophyll binding, oxygen-evolving complex *etc.* Some MF related terms

were observed as upregulated GO terms such as transcription regulator activity, DNA binding transcription factor activity, and molecular function regulator.

KEGG annotation and corresponding GO terms were also identified. In which, “Metabolic pathways (88 gene counts)” and “biosynthesis of secondary metabolites (52)” were upregulated pathways in tropical cultivar along with other 13 pathways as presented in Fig. 10a. Some of the other important KEGG pathways present in the upregulated gene sets of tropical cultivars were important plant pathogen interaction genes, phenylpropanoid biosynthesis, cell wall organising genes (Xyloglucan *etc.*) In addition to various signalling pathways (MAPK signalling and ABC transporters *etc.*) were also identified. Correspondingly, for tropical down regulated genes, “metabolic pathways (78 gene counts)” was the major pathway followed by biosynthesis of secondary metabolites (37) and photosynthesis (25) along with 6 other KEGG pathways such as plant hormone signal transduction, carbon metabolism, glycolysis *etc.* (Fig. 10a and 10b). The enrichment tree depicts the pathways interrelationships for the identified terms for the biological point of view and the terms were similar as explained for the Fig. 10b.

4.8 Gene set enrichment analyses (GSEA)

Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.*, 2005) is a widely used technique that decides whether a set of genes is enriched in a list of differentially expressed genes. To understand the biological context of DEGs, pathway enrichment analysis paves the way. Gene Set Enrichment (GSE) analysis translates information from gene expression profiling into a pathway summary. DE genes are quite repetitively involved in the indistinguishable biological pathways. GSE results intend greater biological interpretability over individual gene analysis. Major categories of gene sets can be selected to customize the search. Furthermore, extends the current GSE methods to RNA-seq data, and provides increased power to detect subtle pathway activity changes, and constitutes a starting point to build pathway centric models of system biology.



*Upregulated genes were not enriched for any of the cellular component GO terms in iDEP, but DAVID identified few terms

Fig. 8a: Gene Ontology terms identified for cellular component (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value.

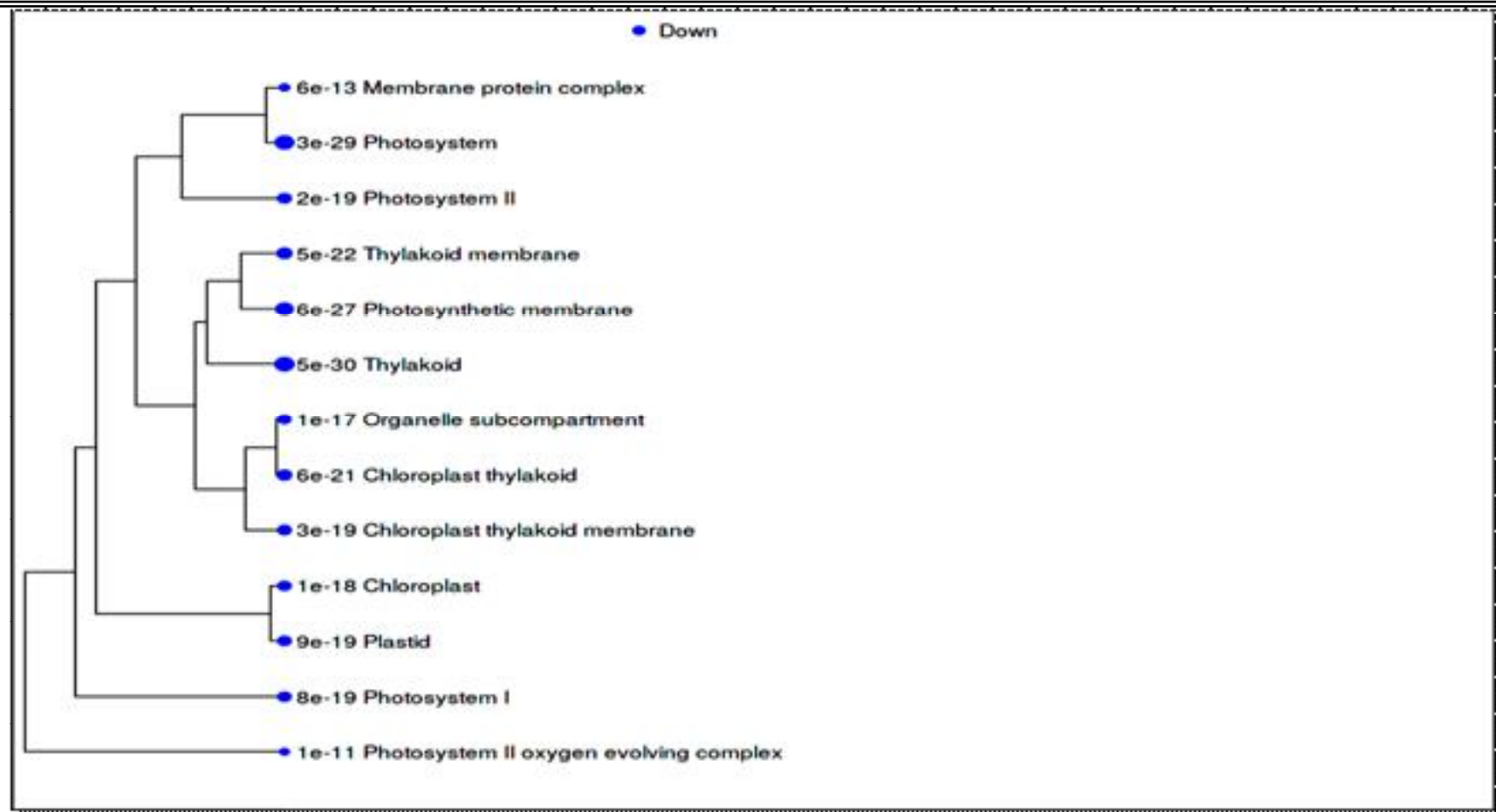


Fig. 8b: Enrichment tree showing relationship of Gene Ontology terms identified for Cellular components (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms ((iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, and brighter blue is highly down regulated) and the size of the terms represents the number of genes that are significant from our list.

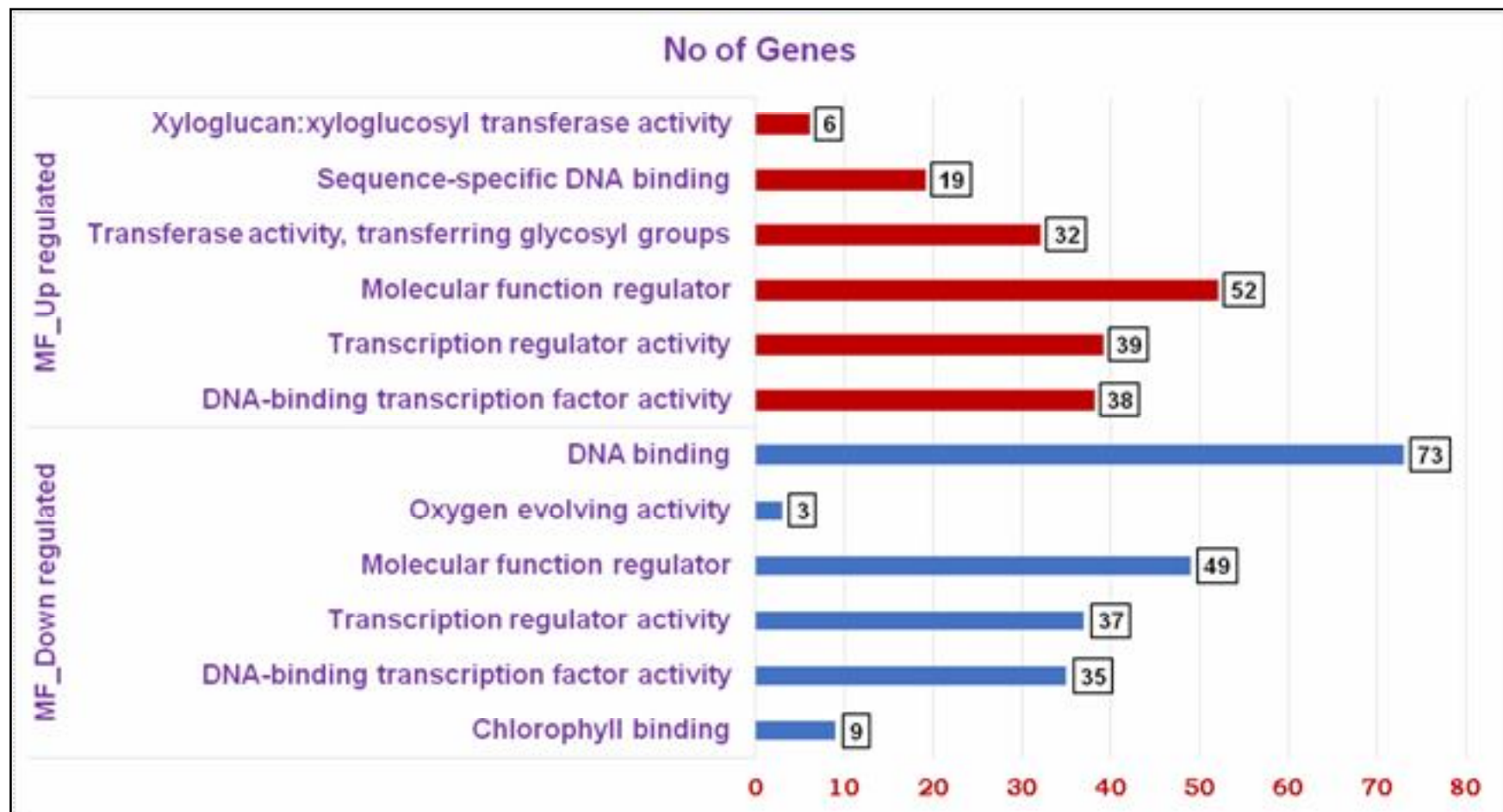


Fig. 9a: Gene Ontology terms identified for ‘Molecular Functions (MF)’ associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value

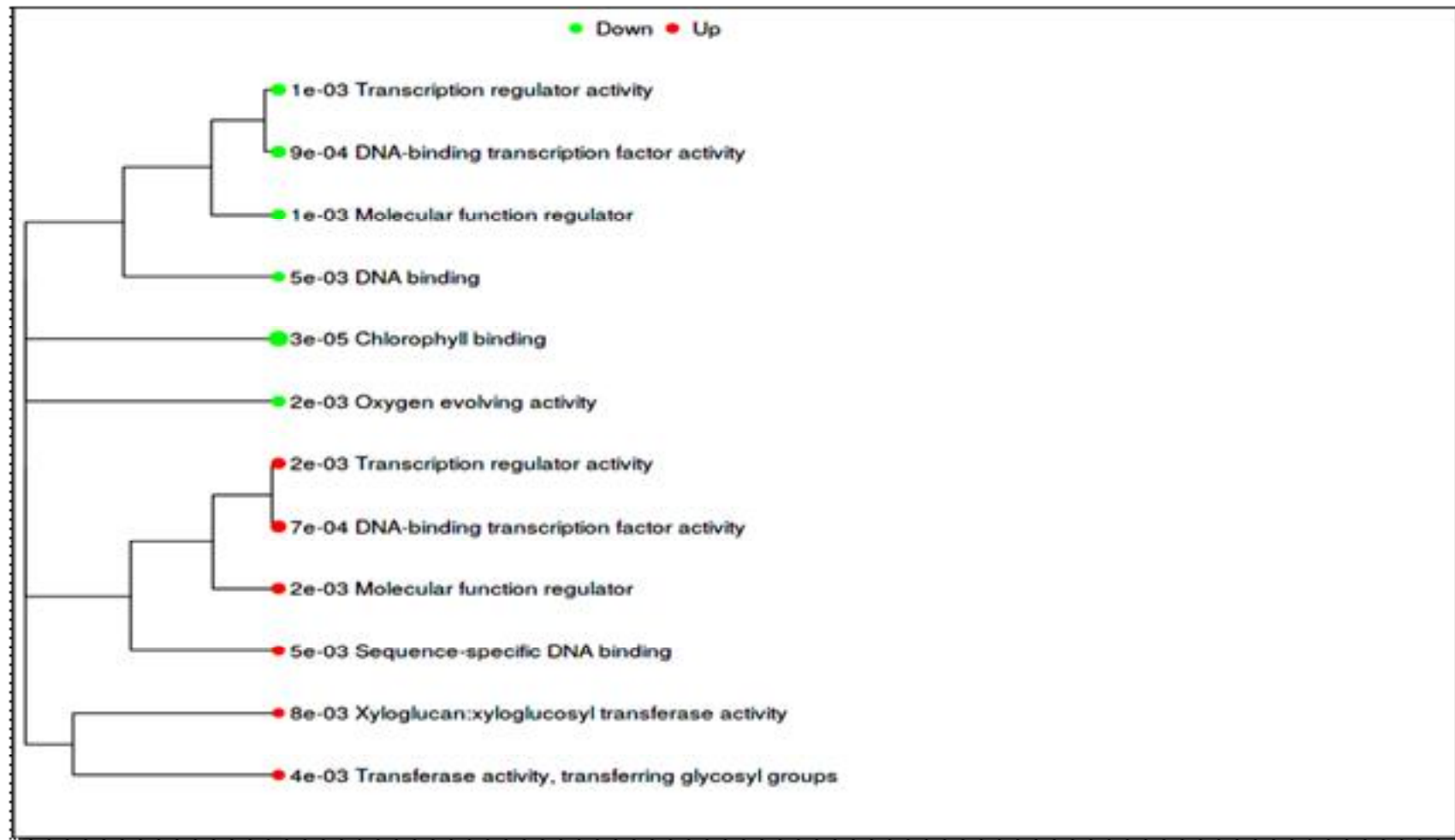


Fig. 9b: Enrichment tree showing relationship of Gene Ontology terms identified for Molecular Functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms (iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, brighter green is highly down regulated) and the size of the terms represents the number of genes that are significant from our list

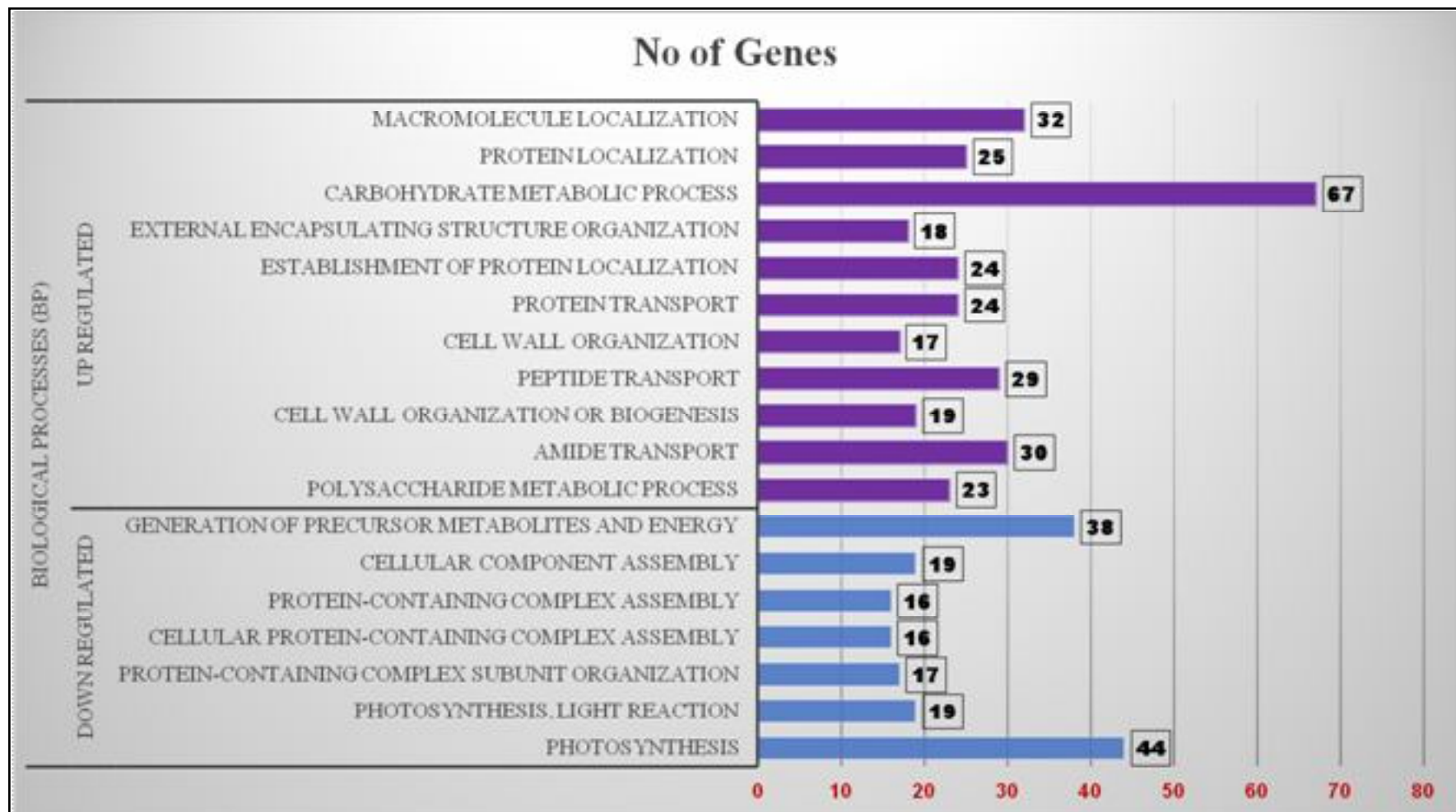


Fig. 10a: Gene set enrichment analysis (GSEA) pathway terms identified for biological process (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value



Fig. 10b: Pathway tree showing relationship of GSEA terms identified for Biological processes (BP) associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms (iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, brighter green is highly down regulated) and the size of the terms represents the number of genes that are significant from our list

Pre ranked FGSEA (fast gene set enrichment analysis) was followed for the pathway enrichment analysis and if a gene set falls at either the top (over-expressed) or bottom (under-expressed) of the list it is said to be enriched. FGSEA is a faster implementation of the GSEA method. FGSEA requires a ranked list of genes and some gene sets to test. There are several ways we could choose to rank our genes; in the present study we could rank by log-fold change (most up-regulated to most down-regulated).

Gene set enrichment analysis (GSEA) pathway terms identified for biological process (BP), Cellular Component (CC) and Molecular function (MF) enrichments associated with tropical carrot root tissue in comparison to temperate carrot root using set of up and down regulated DEGs and P value associated with corresponding GO terms for BP, CC, MF, enrichments is presented in Fig. 10a, 11a, 12a respectively. Pathway tree showing relationship of GSEA terms identified for BP, CC, MF is presented in Fig. 10b, 11b, 12b respectively.

GO terms identified for KEGG enrichment analysis associated with tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID) was also identified. DEGs involved in various metabolic processes, photosynthesis, biosynthesis of secondary metabolites, plant hormone signal transduction and *etc.* are found to be down-regulated in tropical adapted/Asiatic cultivars analogous to temperate adapted/Western cultivars having over expressed/up-regulated pathways for metabolic pathways biosynthesis of secondary metabolites, and carbon fixation in photosynthetic organs as well as carbon metabolism (Table 17). Set of up and down regulated DEGs and corresponding number of the genes associated with the GO terms is presented in Fig. 13a.

However, GSEA terms identified for tropical adapted/Asiatic cultivar are also enriched with metabolic pathways, biosynthesis of secondary metabolites along with important plant pathogen interaction genes, phenylpropanoid biosynthesis in addition to cell wall organising genes (Xyloglucan *etc.*) and various signalling pathways (MAPK signalling and ABC transporters *etc.*) (Fig. 13b).

Table 17: GO term enriched in root transcriptome for BP, CC, MF and KEGG pathway along with NES and gene counts for root transcriptome

S No	A_C ATE	Direction	GO Terms	NES	No of Genes	adj.Pval
1	Biological Processes (BP)	Down Regulated	Photosynthesis	-3.2681	44	1.60E-03
2			Photosynthesis, light reaction	-2.7011	19	1.60E-03
3			Protein-containing complex subunit	-2.274	17	1.10E-02
4			Cellular protein-containing complex assembly	-2.1834	16	1.60E-02
5			Protein-containing complex assembly	-2.1834	16	1.60E-02
6			Cellular component assembly	-2.1183	19	2.40E-02
7			Generation of precursor metabolites and	-2.0149	38	2.60E-02
1		Up Regulated	Polysaccharide metabolic process	2.1876	23	1.60E-02
2			Amide transport	1.9387	30	5.40E-02
3			Cell wall organization or biogenesis	1.93	19	7.70E-02
4			Peptide transport	1.9147	29	6.80E-02
5			Cell wall organization	1.7591		1.90E-01
6			Protein transport	1.7383	24	1.90E-01
7			Establishment of protein localization	1.7383	24	1.90E-01
8			External encapsulating structure organization	1.6984	18	2.20E-01
9			Carbohydrate metabolic process	1.6875	67	1.30E-01
10			Protein localization	1.6857	25	2.10E-01
11			Macromolecule localization	1.6779	32	2.10E-01
1	Cellular Component (CC)	Down	Photosynthetic membrane	-3.2362	47	5.30E-05
2			Thylakoid	-3.1957	57	5.30E-05
3			Thylakoid membrane	-3.1721	41	5.30E-05
4			Chloroplast thylakoid membrane	-3.1294	36	5.30E-05
5			Plastid thylakoid membrane	-3.1294	36	5.30E-05
6			Photosystem	-3.0798	35	5.30E-05
7			Chloroplast thylakoid	-3.0692	40	5.30E-05
8			Plastid thylakoid	-3.0692	40	5.30E-05
9			Organelle subcompartment	-2.9884	43	5.30E-05
10			Membrane protein complex	-2.9847	42	5.30E-05
11			Photosystem I	-2.958	19	5.30E-05
12			Photosystem II	-2.7545	25	5.30E-05
13			Chloroplast	-2.715	83	5.30E-05
14			Plastid	-2.7054	87	5.30E-05
15			Protein-containing complex	-2.5073	85	5.30E-05
16			Oxidoreductase complex	-1.7968	16	3.00E-02
17			Cytoplasm	-1.7285	203	1.90E-03
18			Catalytic complex	-1.5484	25	9.40E-02
19			Plastid stroma	-1.4119	17	1.50E-01
20			Chloroplast stroma	-1.4119	17	1.50E-01
21			Non-membrane-bounded organelle	-1.3692	44	1.50E-01
22			Intracellular non-membrane-bounded organelle	-1.3692	44	1.50E-01
23			Mitochondrion	-1.2855	20	2.30E-01
1		Up	Bounding membrane of organelle	1.5859	17	9.30E-02
2			Cell periphery	1.5325	62	7.20E-02
3			Extracellular region	1.5176	30	1.00E-01
4			Endomembrane system	1.3916	37	1.50E-01
5			Cell wall	1.3835	18	1.60E-01
6			External encapsulating structure	1.3835	18	1.60E-01
1	Molecular Function (MF)	Down	Tetrapyrrole binding	-1.9325	28	1.10E-01
1		Up	Primary active transmembrane transporter	1.885	16	1.10E-01
2			ATPase-coupled transmembrane transporter	1.8222	15	1.10E-01
3			Ubiquitin-protein transferase activity	1.8146	15	1.10E-01
4			Ubiquitin-like protein transferase activity	1.8146	15	1.10E-01
5			Sequence-specific DNA binding	1.8032	31	1.10E-01
6			Hydrolase activity	1.5593	154	1.10E-01
7			Catalytic activity, acting on a protein	1.4975	113	2.20E-01
1	KEGG Pathway	Down	Photosynthesis	-2.875	25	1.80E-04
1		Up	Plant-pathogen interaction	1.7518	20	8.00E-02

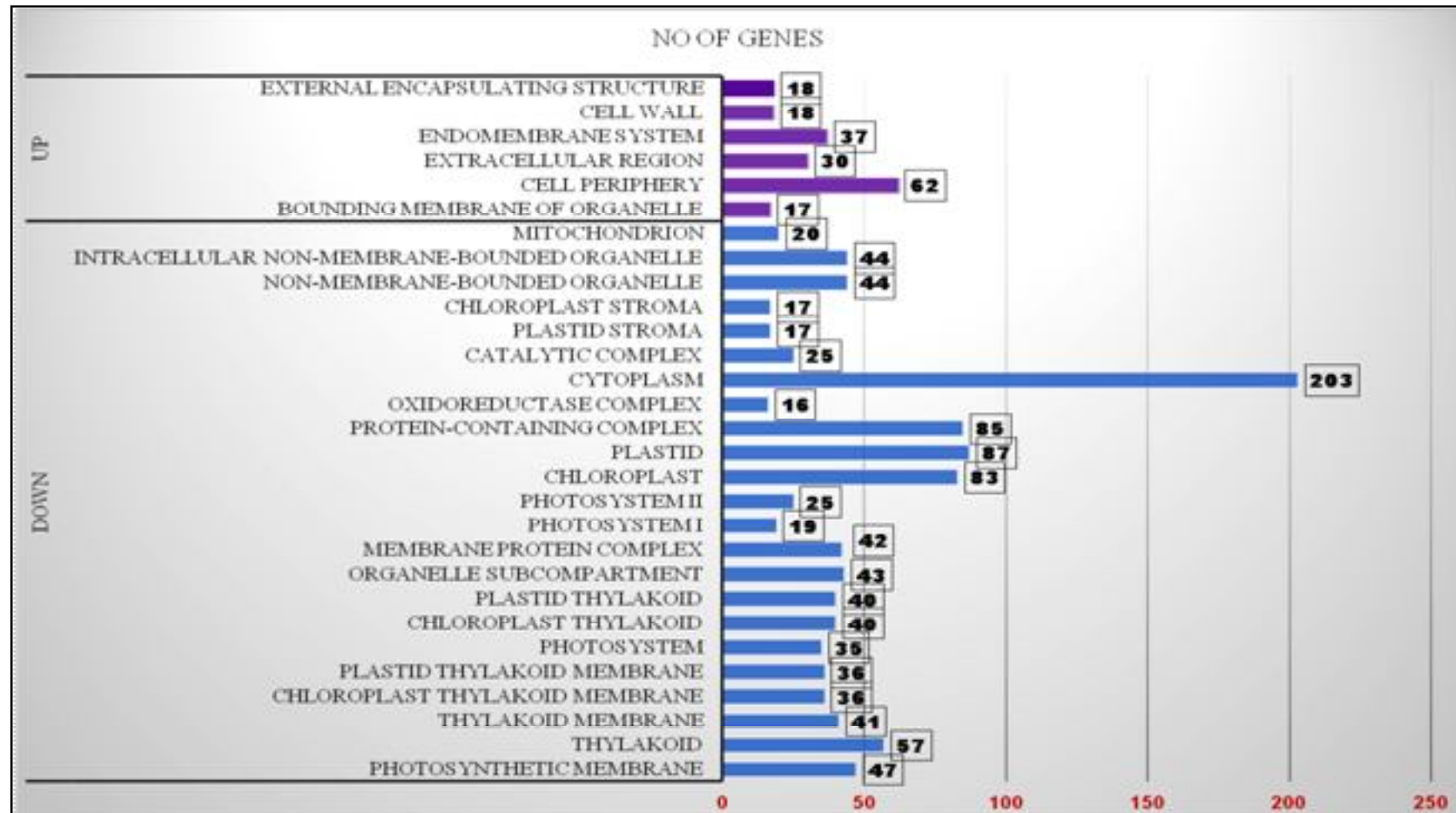


Fig. 11a: Gene set enrichment analysis (GSEA) pathway terms identified for cellular component (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID).). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value

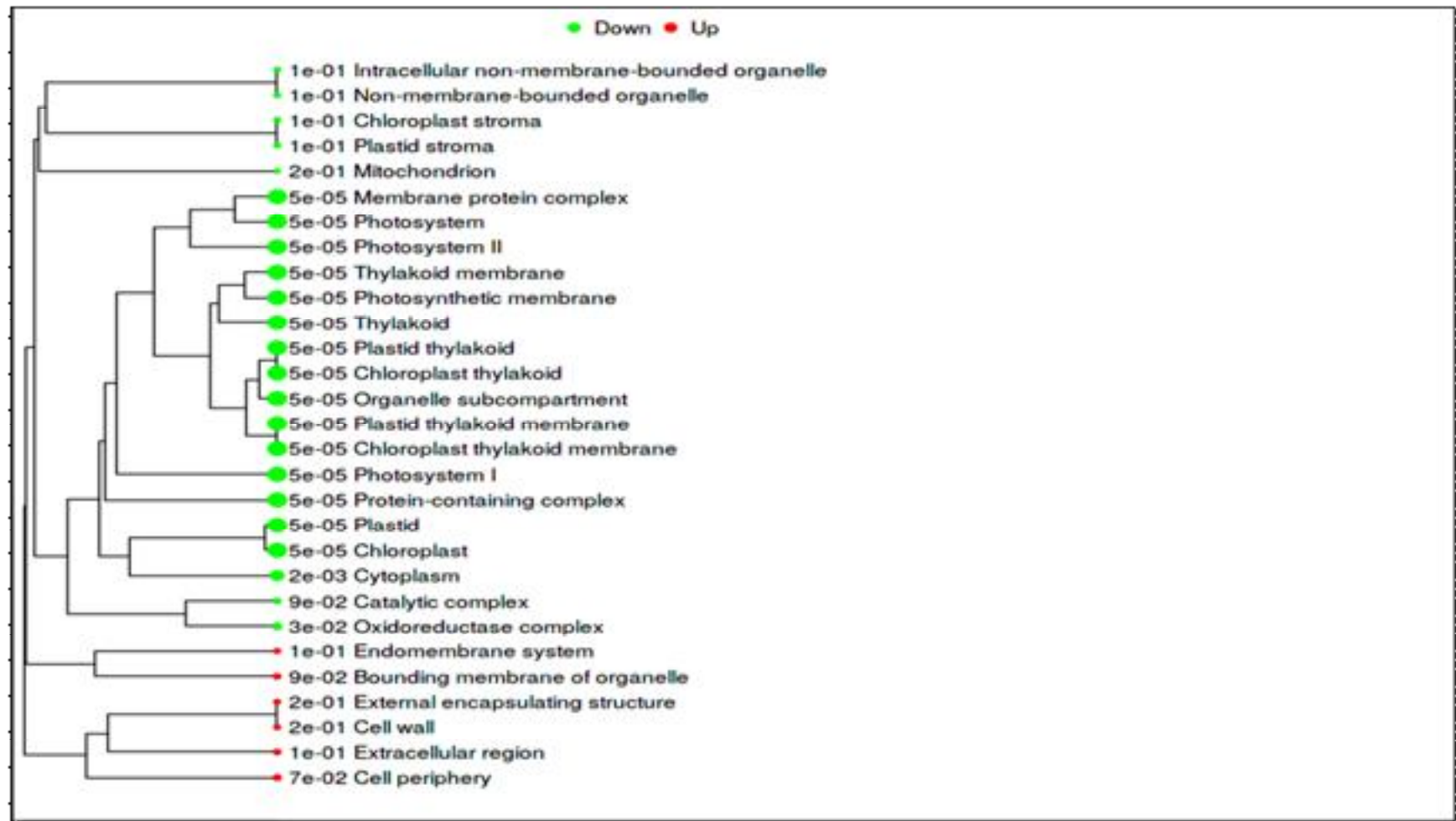


Fig. 11b: Pathway tree showing relationship of GSEA terms identified for cellular components (CC) associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms (iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, brighter green is down regulated) and the size of the terms represents the number of genes that are significant from our list

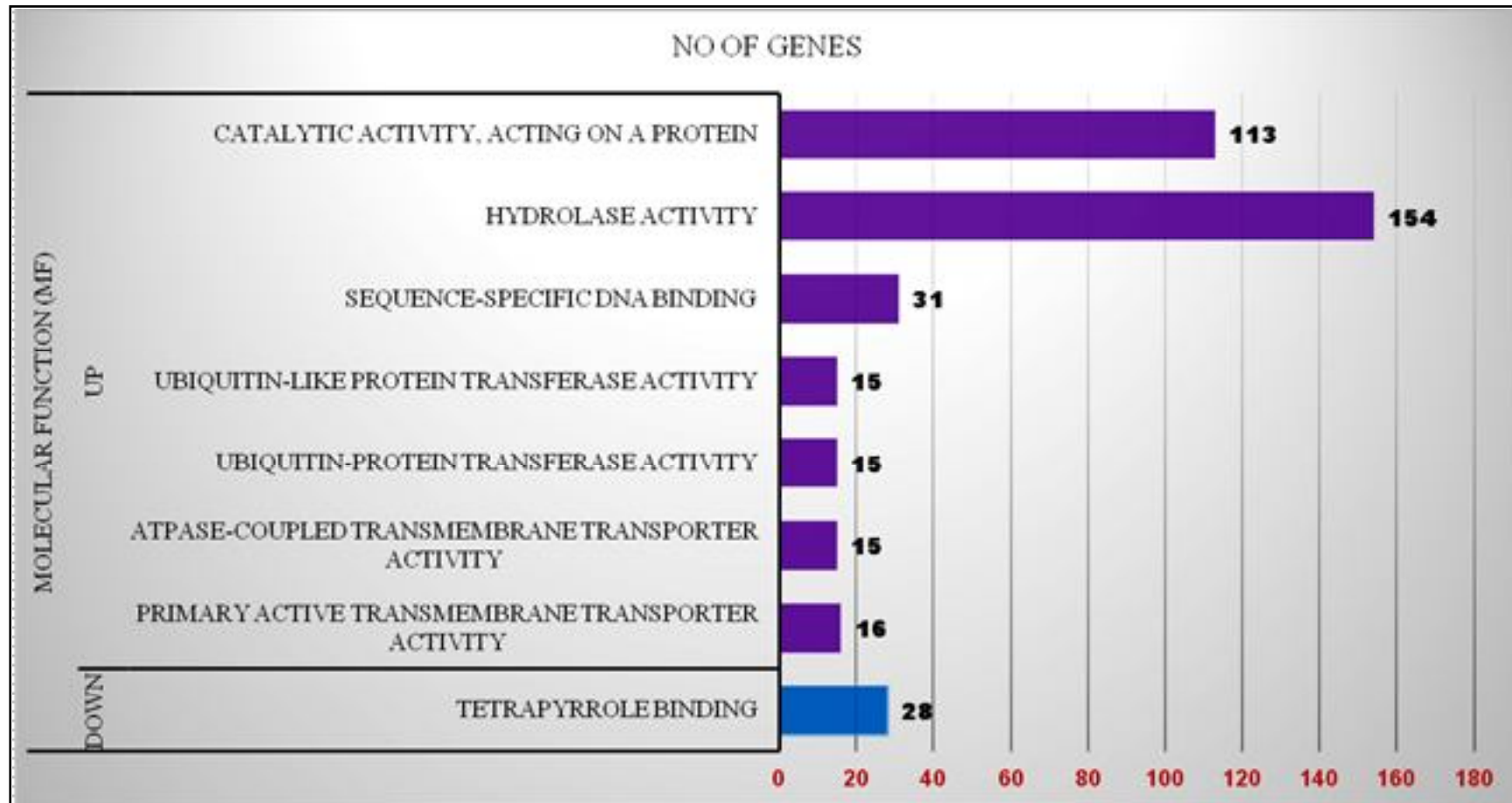


Fig. 12a: Gene set enrichment analysis (GSEA) pathway terms identified for molecular functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root (iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value

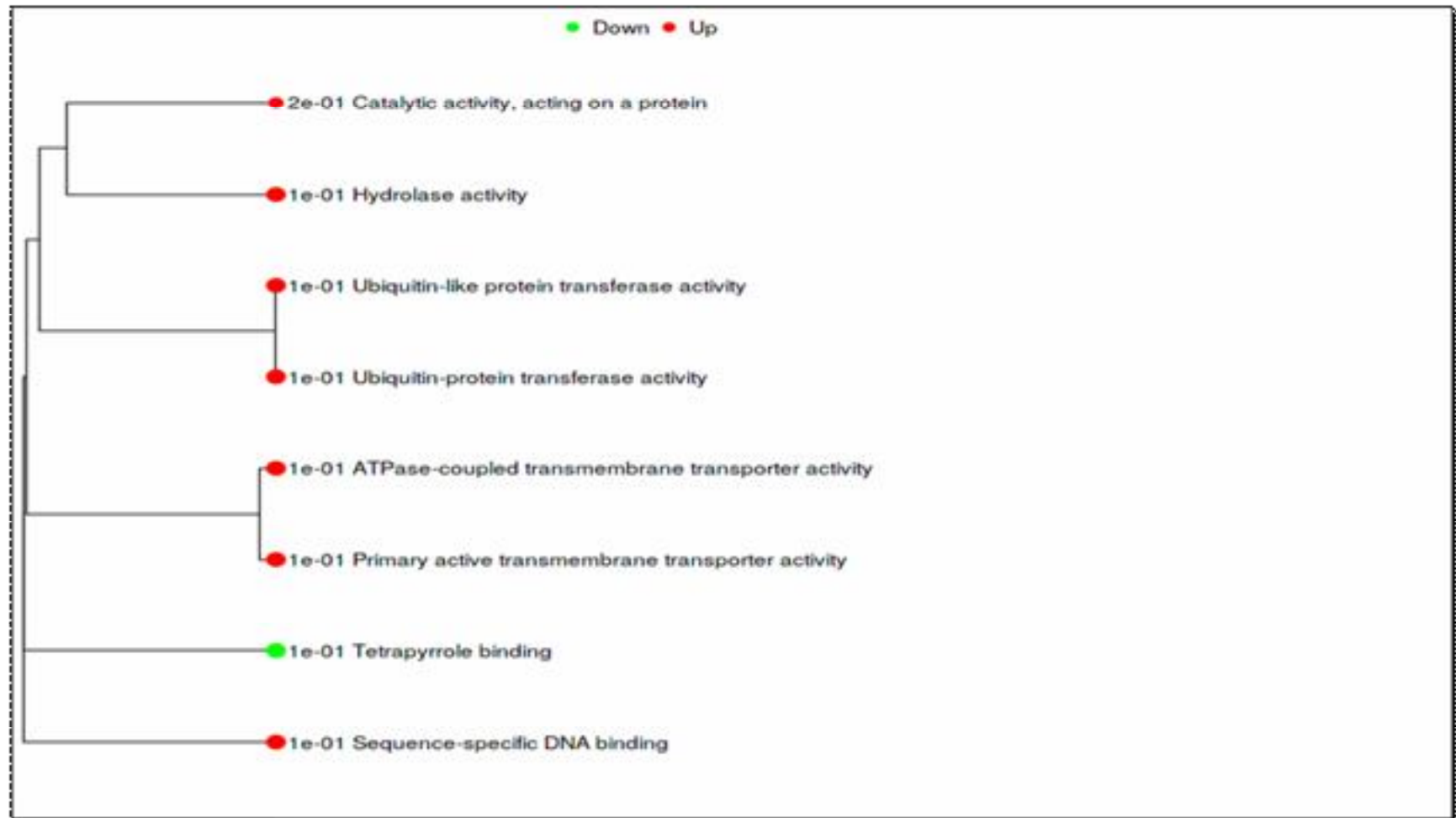


Fig. 12b: Pathway tree showing relationship of GSEA terms identified for Molecular functions (MF) associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms (iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, brighter green is down regulated) and the size of the terms represents the number of genes that are significant from our list

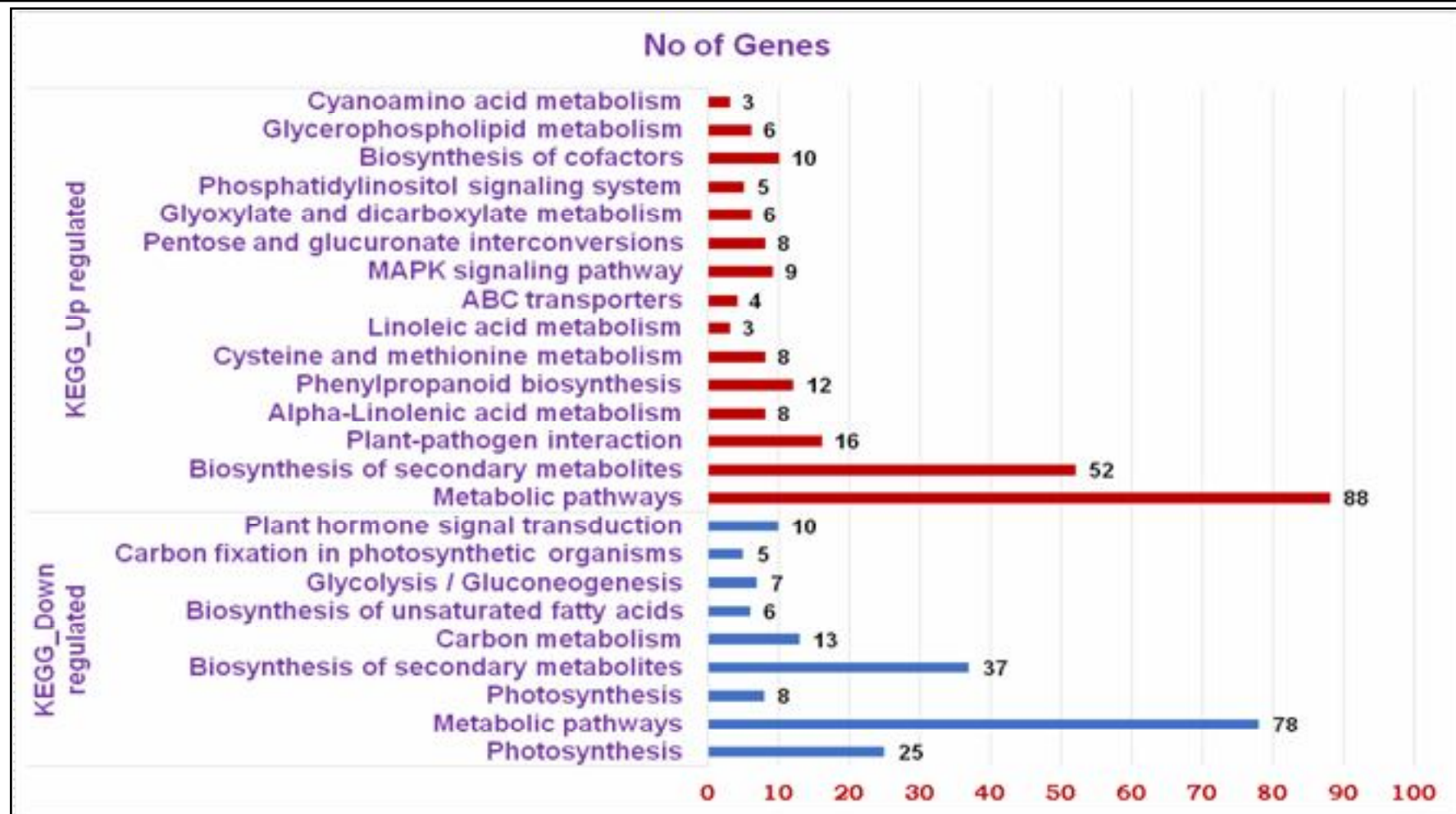


Fig. 13a: Gene Ontology terms identified for KEGG enrichment analysis associated with Tropical carrot root tissue in comparison to temperate carrot root ((iDEP 0.93 and DAVID). Set of up and down regulated DEG and corresponding gene numbers are presented with bar chart. Set of up and down regulated DEG and corresponding number of the genes associated with the GO terms. This plot displays gene ratio (# genes related to GO term / total number of significant genes), not p-adjusted value

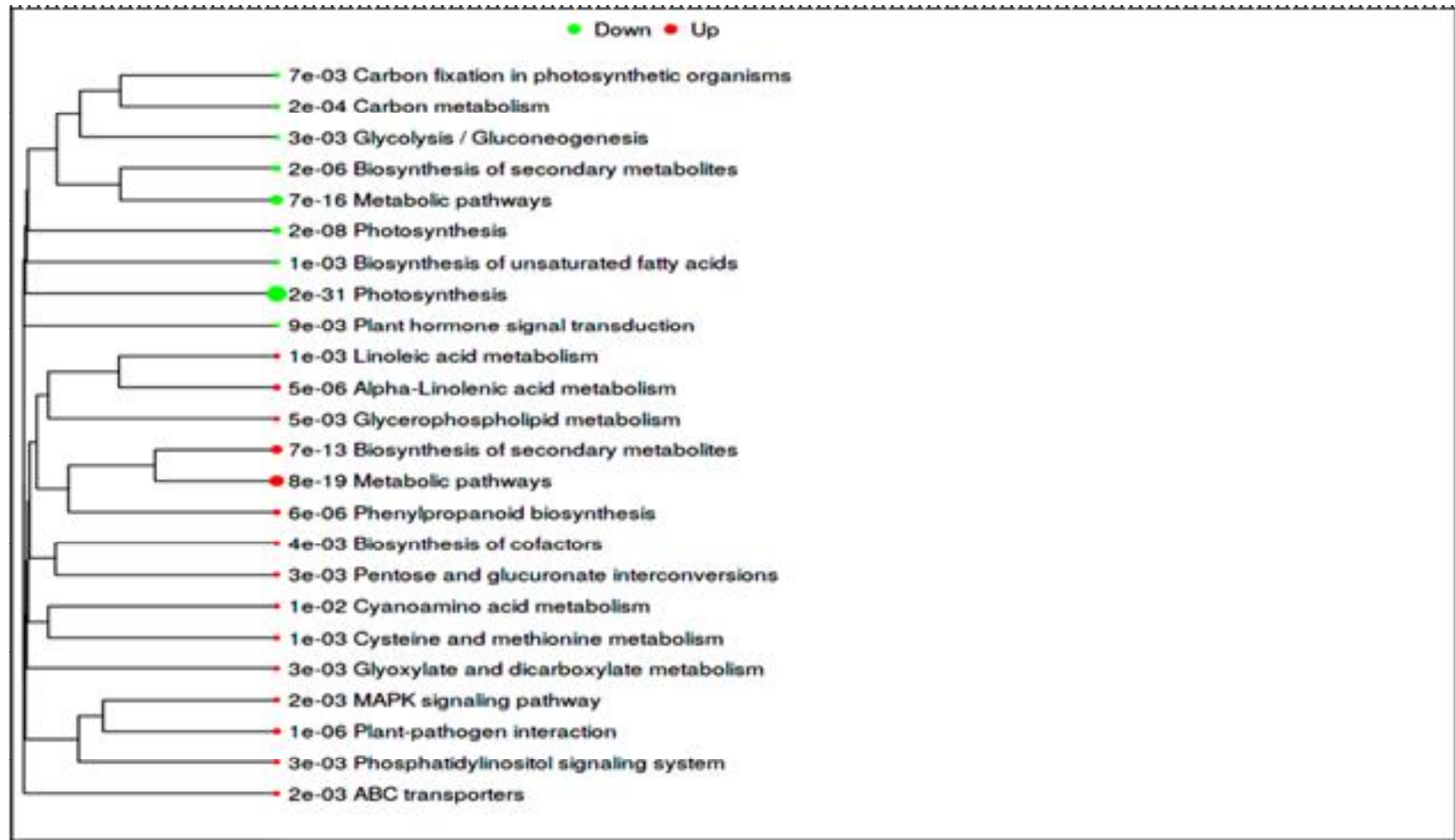


Fig. 13b: Enrichment tree showing relationship of pathways identified for KEGG enrichment analysis associated with Tropical carrot root tissue in comparison to temperate carrot root. A value indicates the P value associated with corresponding GO terms ((iDEP 0.93). The Dot color represents the p-values relative to the other displayed terms (brighter red is more significant and upregulated, brighter green is highly down regulated) and the size of the terms represents the number of genes that are significant from our list

4.9 *De novo* assembly

4.9.1a Quality assessment of assembled root transcriptome

Even though a reference genome is available, it is still recommended to complement a gene expression study by a *de novo* transcriptome assembly to identify transcripts that have been missed by the genome assembly process or are just not appropriately annotated (Hölzer, and Marz 2019). Thus, although a reference genome and several transcriptomes of carrot are available and the previous transcriptome assembly generated by Machaj *et al.*, 2018, compared between wild and cultivated carrots. In the present study, transcripts were assembled by *de novo* method as well to decrease the possibility of losing novel or rare transcripts and cultivar-specific sequences between two important cultivar types of carrot *i.e.*, Asiatic/tropical-type and Western/temperate type cultivars.

Six cDNA libraries were generated from the storage root at vegetative stage of both Asiatic/tropical-type and Western/temperate-type cultivars using Hi-Seq Illumina-4000 deep sequencing. Raw paired end reads (2x100 bp) were obtained from two cultivars with three biological replicates each comprising of six paired end libraries. After removing the low-quality reads and adapters sequences, 130618433 (92.71%) of high-quality cleaned reads were retained for downstream analysis (Table 13). First, the Trinity pipeline v.2.8.6 (Zhao *et al.*, 2011) was used with a minimum k-mer covariance of 2. All the high-quality reads from these six libraries were mixed to generate *de novo* assembled transcriptome. Total 1, 89,857 transcripts were assembled with 1, 03,143 trinity genes by Trinity pipeline (Table 18). A mean contig length of 1093.67 bp and an N50 of 1890 bp and ExN90 1950 bp, were generated which is comparable to similar RNA-seq reports in carrot. After clustering with CD-HIT-EST, the contigs were assembled into 44,856 transcripts from root transcriptome. The combined transcriptome assembly of tropical and temperate cultivars contains average contig lengths 472 bp and 14,803 contigs are equal or longer than 1000 bp. Finally, we identified unigenes in non-redundant combined transcriptome assemblies by blasting against the carrot reference genome. This analysis resulted in 44,856 unigenes in combined assembly of tropical and temperate root transcriptome assembly. Thus, we have successfully constructed a good quality assembly from Illumina paired-end sequencing of root transcriptome. The length of the contigs present in *de novo* assembly ranged from 189 to 26545bp (Fig. 14a).

4.9.1b Quality assessment of *De novo* assembled FPL transcriptome

To obtain an overview of the flowering pathway between tropical (tropical) and temperate (temperate) cultivars a transcriptome, RNA-Seq libraries were prepared from pooling of RNA from Floral primordia and leaf (FPL) tissues at vegetative to reproductive phase transition stage with three biological replicates for each tropical and temperate genotype. A total of 233.29 million high quality 151 bp paired-end reads were generated. The good quality reads were *de novo* assembled (202.36 M). Total 1,23,912 transcripts were assembled with 49978 trinity genes by Trinity pipeline with K-mer size of 25 (Table 18). A mean contig length of 972.8 bp and an N50 of 1766 bp and ExN90 1925bp, were generated which is comparable to similar RNA-seq reports in carrot. After clustering with CD-HIT-EST, the contigs were assembled into 49,978 transcripts from FPL transcriptome and are referred as unigenes. The combined transcriptome assembly of tropical and temperate cultivars contains average contig lengths 972.80bp and 55,899 contigs are equal or longer than 1000 bp in the trinity assembly. Whereas, in the unigenes assembly, average contig length of 841.76 bp and 13694 contigs are equal or longer than 1000 bp were obtained. The length of the contigs ranged from 180bp to 26995bp in trinity assembly and 195bp to 16797 bp lowest to highest contig length among the unigenes were observed (Fig. 14b).

4.9.1c Assembly quality assessment and normalization

The quality of the assemblies was assessed using several methods. First, the assemblies were evaluated against a database of single copy orthologous genes for plants as implemented in BUSCO (version 1.161). The completeness of the *de novo* assemblies generated for root and FPL transcriptome was evaluated through Benchmarking Universal Single-Copy Orthologs (BUSCO). BUSCO analysis indicates the number of conserved orthologs presents in transcriptome, determined percentage of mapped reads, properly paired reads and singleton. The completeness of gene sets and transcriptomes was identified using their gene content as a complementary method beside evolutionarily conserved orthologous gene groups obtained from OrthoDB. After BUSCO assessment; secondly, the processed RNA-seq reads were mapped to the own transcriptome assembly to determine percentage of mapped reads, properly paired reads and singleton. The quality of assembled transcriptome details is discussed as below.

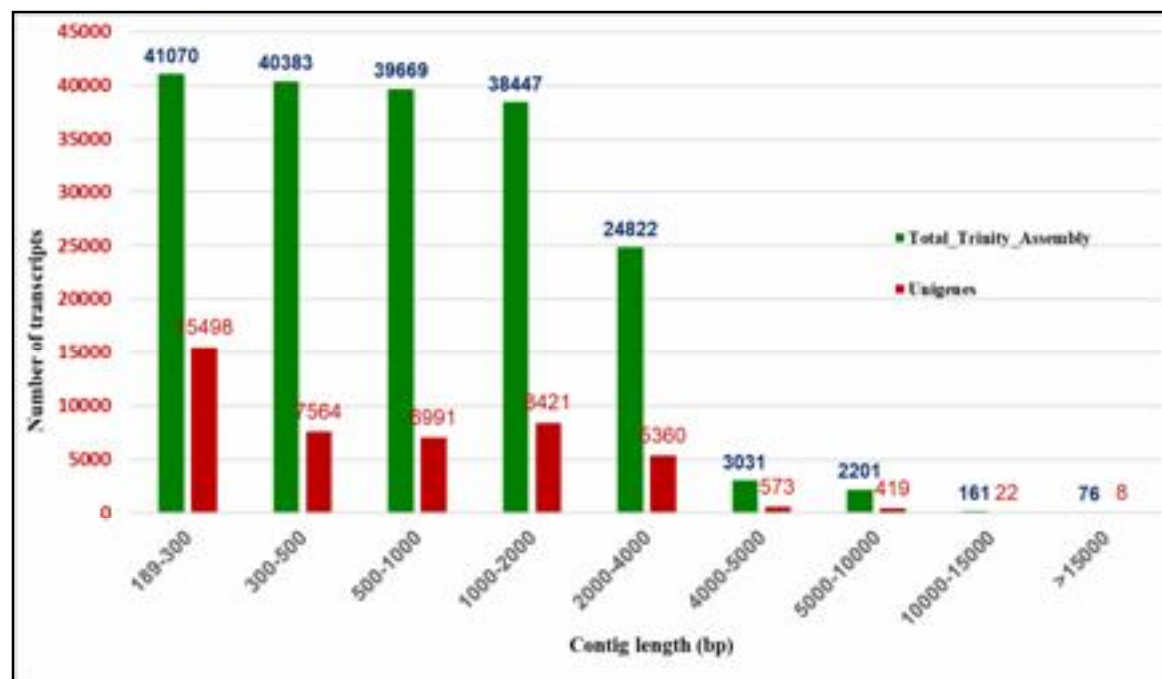


Fig. 14a: Comparison of contig lengths (bp) between trinity assembly and unigenes in *de novo* assembled root transcriptome of tropical and temperate cultivars

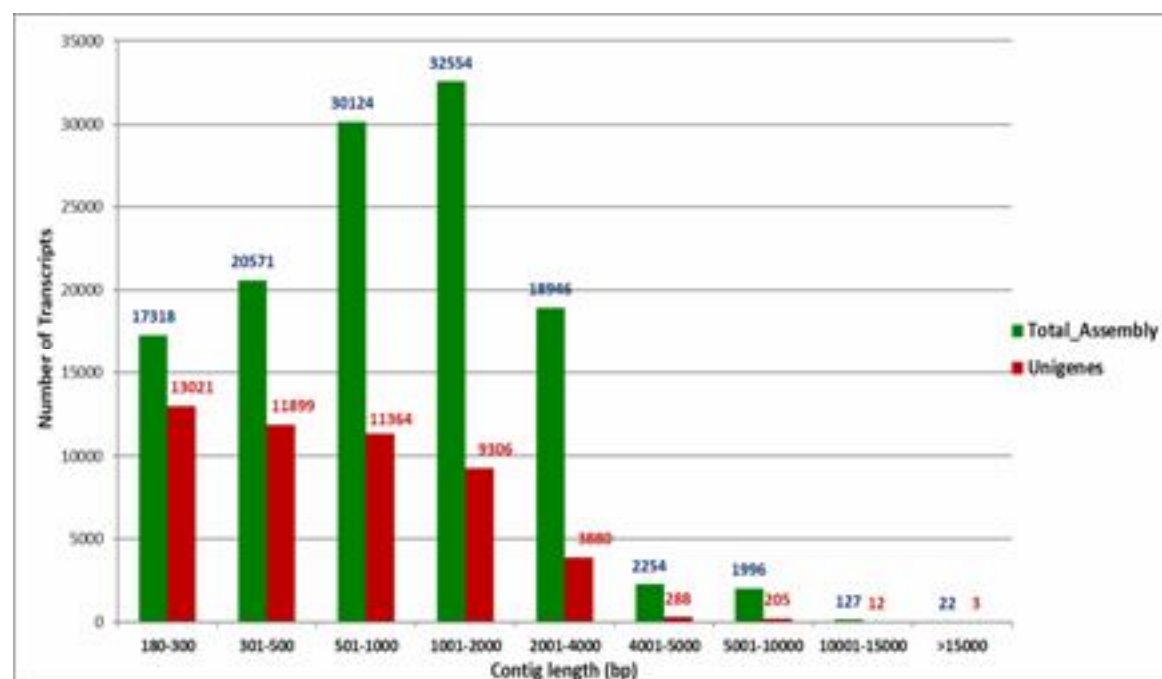


Fig. 14b: Comparison of contig lengths (bp) between trinity assembly and unigenes in *de novo* assembled FPL transcriptome of tropical and temperate cultivars

Table 18: Details of Denovo Assembly and unigenes identified after CD-HIT_EST in Root and FPL *de novo* assembly

Trinity Assembly output	Root transcriptome (Vegetative phase)		FPL tissue transcriptome (Reproductive phase)	
	Total Assembly Output	Unigenes (CD_HIT Longest isoform)	Total Assembly Output	Unigenes (CD_HIT Longest isoform)
Total assembled bases	20,76,41,471	4,39,61,967	50,174,025	42069522
Total transcripts	1,89,857	44,856	1,23,912	49978
Total 'genes'	1,03,143	44,856	51,577	49978
Per cent GC	39.59	40.02	39.38	39.70
Contig N50 (bp)	1890	1843	1766	1363
Median contig length (bp)	623	472	509	502
Average contig (bp)	1093.67	980.07	972.8	841.76
ExN90*	1950 bp (3967)	2086bp (4696)	1925bp (8174)	1813bp (7929)

* Numbers in the parenthesis indicates the transcript numbers

A) BUSCO analyses of root transcriptome

BUSCO completeness estimations showed that 89.08% of the BUSCO groups have complete gene representation (single-copy or duplicated), while 2.84% are only partially recovered, and 8.08% are missing. BUSCO assessment of the quality of transcriptome assemblies revealed transcriptome assembly with high number of gene representation (2326 BUSCOs), high completeness, 89.08% (47.81% complete single copy + 41.27% Complete Duplicated), low percentage of fragmented 2.84% and low missing BUSCOs (8.08%) presented in Table 19. The BUSCO genes were selected among groups of orthologs selected from eudicots_odb10 lineage across 31 genomes matching specific evolutionary expectations across set of orthologs.

B) BUSCO analyses of FPL tissue transcriptome

To assess the completeness of our FPL tissue transcriptome and the level of redundancy, BUSCO analysis was performed which revealed a completeness score of 85.38 % with low percentage of fragmented 4.69 % and low missing BUSCOs (9.93 %) presented in Table 19. This indicates that most of the core gene set is present in our assembly, suggesting a high-quality assembly. However, as anticipated, high level (34.74%) redundancy was evident as expected in carrot.

C) Alignment of RNA-seq reads from individual root and FPL tissues back to respective *de novo* generated assembly

We then aligned our RNA-seq data from individual root libraries back to the assembled transcriptome using Bowtie to assess the quality of assembly. Approximately 99% of the aligned RNA-seq reads of root transcriptome were proper pairs (*i.e.*, each end of a paired-end read successfully mapped to the same contig) as shown in Table 20a. Whereas, FPL transcriptome showed 57.87% to 63.18 % alignment of RNA-seq reads from individual libraries back to unigene assembled transcriptome as presented in Table 20b.

Table 19: Results of the carrot root and FPL tissue transcriptome assembly assessment using BUSCO

BUSCO Notation	Root transcriptome		Flower Transcriptome	
	Percentage	BUSCO Groups	Percentage	BUSCO Groups
BUSCO Completeness	89.08%	2326	85.38 %	2326
Complete Single copy	47.81%	1112	50.64%	1178
Complete Duplicated	41.27%	960	34.74 %	808
Fragmented (Partially recovered)	2.84%	66	4.69 %	109
Missing	8.08%	188	9.93 %	231

Table 20a: Alignment rate of individual library of root transcriptome with Trinity assembly

Alignmen t rate	Temperate _R1	Temperate _R2	Temperate _R3	Tropical_ R1	Tropical_ R2	Tropical_ R3
Total Reads	24,934,401	18,142,740	21,226,283	26,766,083	17,630,817	21,918,109
Aligned concordantly exactly 1 time	21846481 (87.62%)	13888483 (76.55%)	17931542 (84.48%)	22260494 (83.17%)	14200474 (80.54%)	17534373 (80.00%)
Aligned concordantly >1 times	1972282 (7.91%)	3293720 (18.15%)	2094894 (9.87%)	3505825 (13.10%)	2594571 (14.72%)	3100171 (14.14%)
overall alignment rate	99.20%	98.54%	99.03%	99.08%	98.77%	98.66%

Table 20b: Alignment rate of individual FPL tissue transcriptome library with Unigene assembly

Alignment rate	Temperate _R1	Temperate _R2	Temperate _R3	Tropical _R1	Tropical _R2	Tropical _R3
Total Reads	26223523	41939945	38974029	31659119	30948384	32614387
Aligned concordantly exactly 1 time	14306177 (54.55%)	22827453 (54.43%)	22534959 (57.82%)	18267860 (57.70%)	17866740 (57.73%)	17728734 (54.36%)
Aligned concordantly >1 times	869558 (3.32%)	1902883 (4.54%)	944289 (2.42%)	1383999 (4.37%)	1686338 (5.45%)	1193840 (3.66%)
Overall alignment rate	57.87 %	58.97 %	60.24 %	62.07 %	63.18 %	58.02 %

4.9.1d Transcript quantification and normalization using RSEM

The third stage in the pipeline requires obtaining transcripts isoform abundance estimates for each of the samples. The transcript abundance was measured for each cultivar and replicates individually as the expected number of fragments per kilo base (kb) of transcript sequence per million mapped reads (FPKM) using RSEM (RNA-Seq by Expectation-Maximization) version 2.8.6, installed in trinity pipeline (v2.8.6). RSEM is a software tool that estimates gene and isoform expression levels from our paired-end RNA-seq data. Later, the developed matrix from RSEM estimates represented relative transcript abundance in all six libraries.

4.9.1e Transcript's filtering:

To reduce the number of potential spurious contigs, we filtered them based on the abundance and quality of the assembly. For abundance, we considered isoforms expression level and redundancy. Expression levels were detected using the RSEM method, and only highly expressed isoforms were retained for further analysis.

Redundancy was eliminated by clustering our *de novo* assembled transcripts with highly similar contigs using CD-HIT-EST at a nucleotide identity of 95%. Fewer than 3% of the Trinity transcripts were redundant and were therefore removed. The numbers of final non-redundant transcripts were further filtered by subjecting to longest isoform selection. After extracting the longest transcript for each gene, 44,856 unigenes were obtained from root transcriptome. The average length of unigene was 980.07 bp, and approximately 38.88% of the unigenes were at least 500 bp long. N50 of the unigenes was 1843bp Table 18a discussed above.

Whereas, in FPL tissue Trinity assembly generated 1,23,912 transcripts with 51,577 genes. However, after CD_HIT-EST followed by filtering for longest isoform, a total of 49,978 unigenes with average length of 841.76 bp long was obtained from FPL transcriptome. N50 of the unigenes of FPL tissue was 1363 bp (Table 18).

After *de novo* assembly and quality assessments, we performed functional annotations and determined differentially expressed genes (DEGs). Overall, we presented a stringent workflow in this study for processing of RNA-seq data, obtaining

de novo assembly, assessment of assembly, functional annotation and differential gene expression. The expression estimates generated by RSEM from our assembled transcriptome with respect to both cultivar's transcriptome was used with two differential expression analysis tools, edgeR and DESeq to determine which genes are being differentially expressed.

4.10 Identification of differentially expressed genes by *de novo* assembly

The next step is to test for differential expression of genes between the experimental tropical and temperate carrot groups. We begin our differential gene expression analysis by creating a matrix containing the counts of RNA-seq fragments for each sample using the expected count data produced by RSEM. The reads were counted by mapping to assembled *de novo* transcriptomes to identify DEGs between Asiatic type (tropical adapted) and Western type (temperate adapted) root samples.

The fasta file generated from CD-HIT-EST redundancy removal followed by longest isoform retention step, 44,856 unigenes were retained in root transcriptome and 49,978 unigenes were retained in FPL transcriptome. These fasta files consisting of unigenes of root and FPL were once again used for RSEM method of transcript quantification for individual samples of 6 libraries in each root as well FPL tissue separately. RSEM Matrix was once again generated for the six files with only read counts represented from the unigenes. The generated RSEM matrix file was used for DEG identification by DESeq2 and edgeR tools.

Gene expression was estimated by negative binomial distribution to model transcript counts and determines gene-specific variability. edgeR identified the relative expression counts for each gene after the Trimmed Mean of M-values (TMM) normalization method accounting sequencing depth, RNA composition, and gene length. In DESeq gene expression is expressed in terms of Relative Log Expression (RLE) considering sequencing depth and RNA composition.

To identify various gene responses of carrot root transcriptome and FPL transcriptome related to several biological pathways functioning in tropical region and examine DEGs between tropical adapted (Asiatic type) and temperate adapted (Western

type) cultivars a pairwise approach was used in both DESeq and edgeR. The following subsections provide implementation and comprehensive results for the differential expression analysis, for both of the software packages used (edgeR and DESeq).

4.10.1 DEG identification between temperate and tropical adapted cultivars in root transcriptome

Only genes represented with an adjusted P-value lower than 0.01 and at least a two-fold change were considered as significantly differentially expressed in the pairwise comparison of the samples. First, genes with zero counts were eliminated since it makes no sense to test them for differential expression if they were not expressed. Under these criteria, A total of 2616 and 2250 DEGs were identified respectively in DESeq2 and edgeR. This comprised of 1335 and 1152 up-regulated genes, 1282 and 1105 downregulated in DESeq and edgeR respectively (Fig. 15a). In this comparison a larger number of DEGs were identified in DESeq contrasting to edgeR. However, 2239 common DEGs were identified as significantly differentially expressed genes (DEGs) by these two tools. In which, 1149 genes were significantly up-regulated (accounting for 51.31%) in temperate or Western type carrots instead down-regulated in tropical or Asiatic type carrots. Compared to the Asiatic type carrots, 1090 set of genes were downregulated in temperate or Western type (48%) and conversely the same set genes were up-regulated in the tropical or Asiatic type carrots.

Here, onwards we will describe only the 2239 DEGs that were common in root transcriptomes (temperate + tropical) to support the validity and reliability of our analysis and to obtain more relevant data from the two groups.

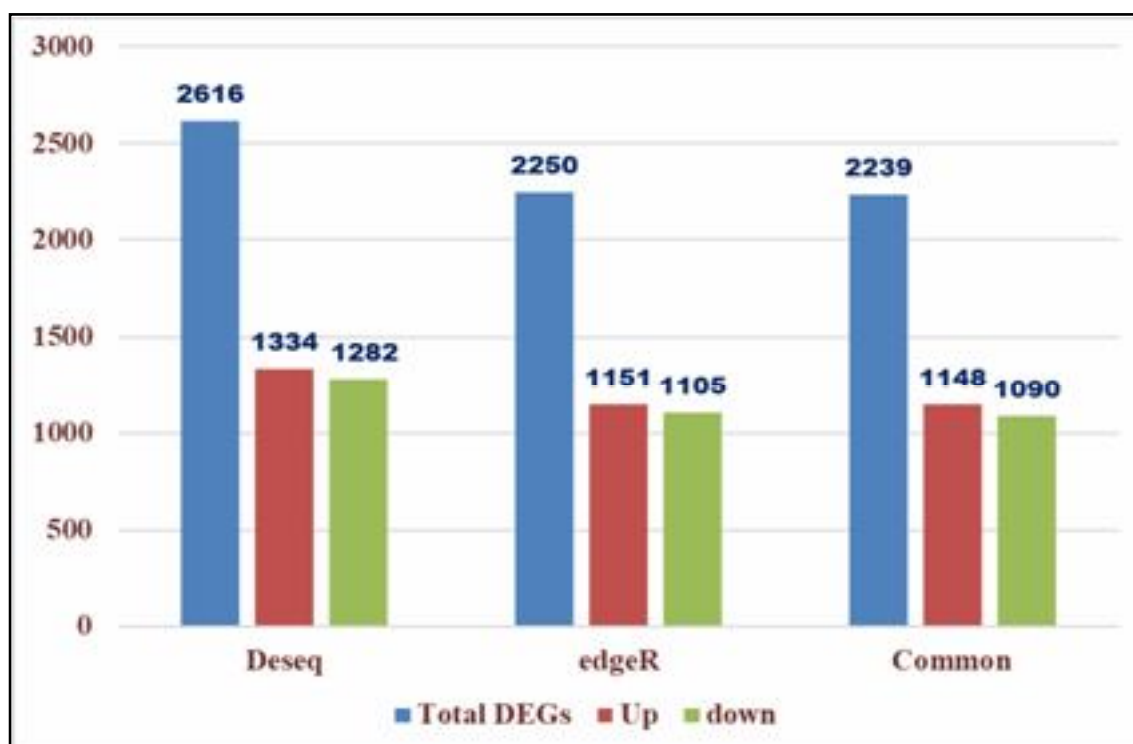
The combinations of p-value and fold-change threshold criteria $\log_2$ fold-change ≥ 2 and $p \leq 0.01$ seem to give more biological meaningful sets of genes than using either of them alone. The top DEG tags have tiny p-values and FDR values, as well as large fold changes. Fig. 15b represents the volcano plot, and significant genes satisfying $p\text{-value} \leq 0.01$ $\log_2\text{FC} \geq 2$, and $\log_2\text{FC} \leq -2$ are labelled and shown in red and blue dots for up and down regulated genes between temperate and tropical root transcriptome respectively.

4.10.2 DEG identification across temperate and tropical adapted cultivars in FPL transcriptome

Reads were counted by mapping to de novo assembled transcriptomes for the identification of differentially expressed genes (DEGs) temperate and tropical adapted FPL transcriptome. The number of DEGs identified in the different comparisons can be found in Fig. 16a. The statistically and biologically significant genes were selected by considering all the below mentioned factors. In the analysis, a particular gene 'i' is defined as differentially expressed either up-regulated or down-regulated, if it satisfied the following set of criteria

- ❖ FDR (p-adjusted) = 0.05; to avoid false positive estimation.
- ❖ $\log_2\text{-foldchange (counts)} > 1.5$ or 2-fold to take into account biological significance

Under this criterion, a total of 1746 with 946 upregulated and 800 downregulated genes were identified in DESeq2. Total of 1794 DEGs including 980 upregulated and 814 downregulated DEGs were identified in edgeR between temperate and tropical FPL transcriptome respectively (Fig. 16a). In this comparison a larger number of DEGs were identified in DESeq contrasting to edgeR. However, 1671 common DEGs were identified as significantly differentially expressed genes (DEGs) by these two tools as shown by in Fig. 16a). In which, 898 genes were significantly up-regulated in temperate cultivar compared to tropical carrots. Whereas, 773 set of genes were downregulated in temperate or up-regulated in the tropical cultivars. However, here onwards, we will describe only the 1671 genes that were common in both edgeR and DESeq across temperate or tropical cultivar FPL transcriptome to support the validity and reliability of our analysis. Fig. 16b represents the volcano plot, and significant genes satisfying $p\text{-value} \leq 0.01$ $\log_2\text{FC} \geq 2$, and $\log_2\text{FC} \leq -2$ are labelled and shown in red and blue dots for up and down regulated genes between temperate and tropical FPL transcriptome respectively.



* Up and down regulated are in the context of temperate (temperate cultivar) over tropical cultivar.

Criteria used for DGE identification p value=0.05, FDR=0.001, Log2FC- Upregulated ≥ 2 and downregulated ≤ -2

Fig. 15a: Summary statistics of DEGs identified by DESeq, edgeR tools and number of common DEGs selected in root tissue across temperate and tropical cultivars

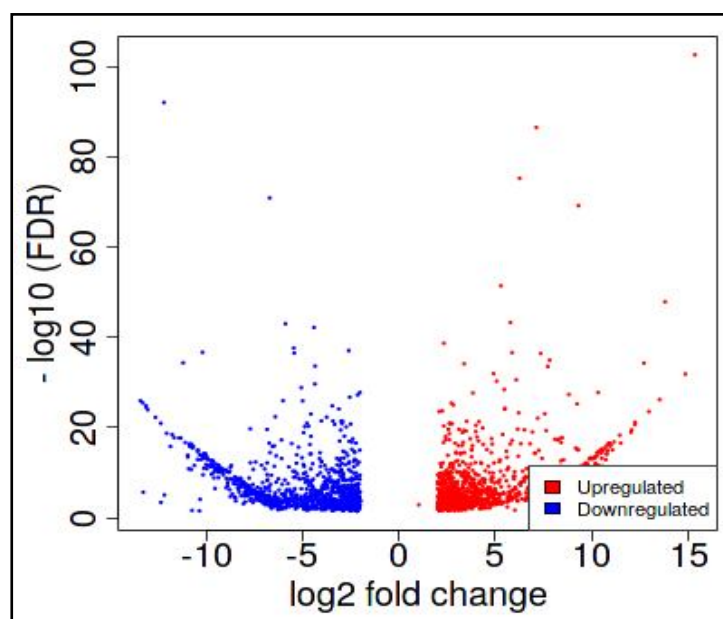
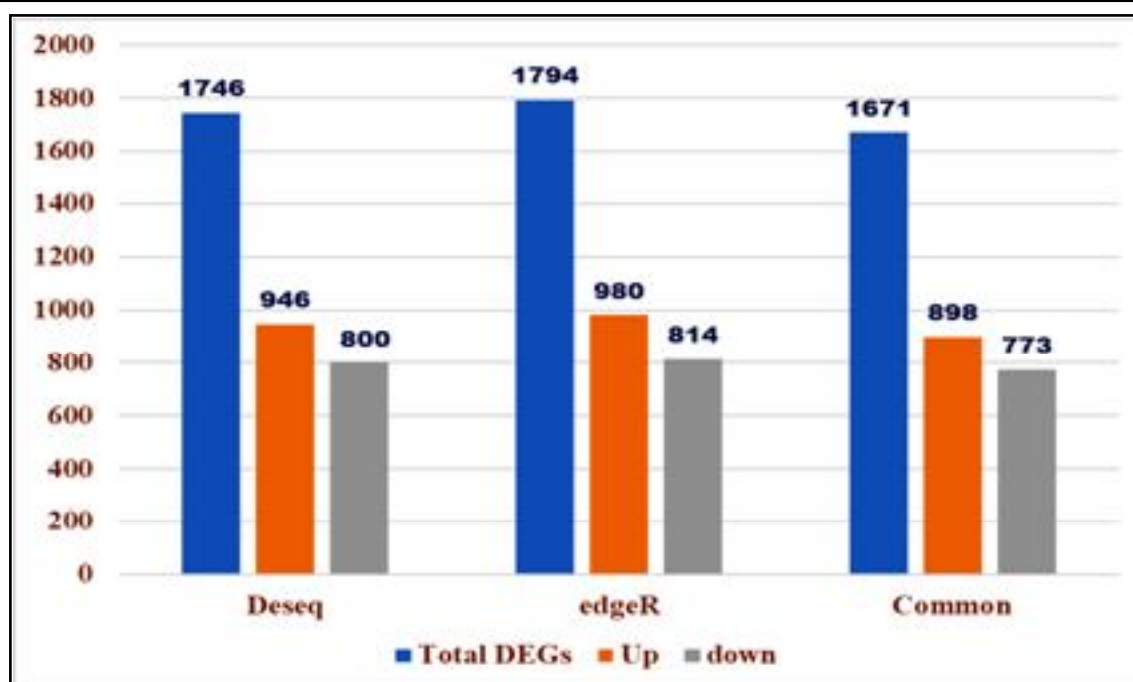


Fig. 15b: Visualization of total DEGs identified by *de novo*-based root transcriptome between Asiatic and Western carrot cultivars by volcano plot



* Up and down regulated are in the context of temperate cultivar over tropical cultivar.

Criteria used for DGE identification p value=0.05, FDR=0.001, Log2FC- Upregulated ≥ 2 and downregulated ≤ -2

Fig. 16a: Summary statistics of DEGs identified by DESeq, edgeR tools and number of common DEGs selected in FPL tissue across temperate and tropical cultivars

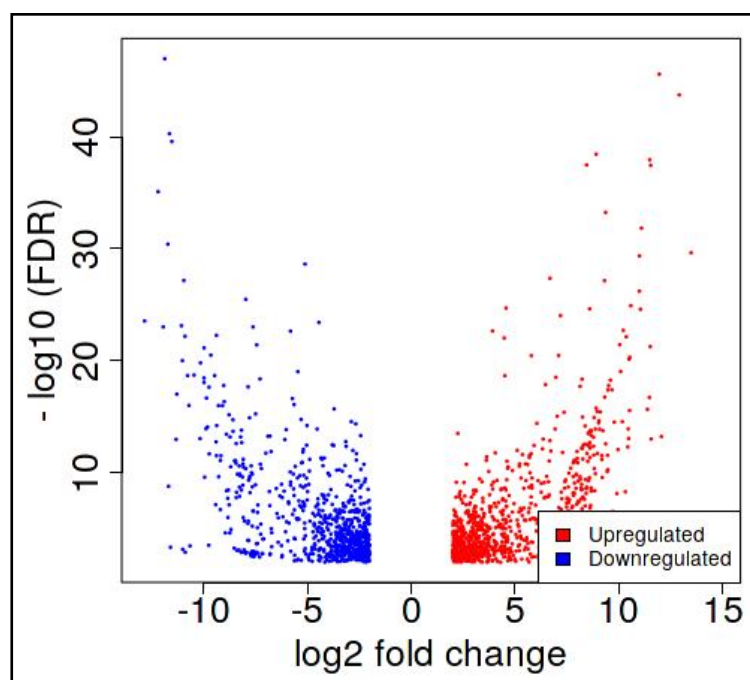


Fig. 16b: Visualization of total DEGs identified by *de novo*-based FPL transcriptome between temperate and tropical carrot cultivars by

4.11 Functional annotation and GO classification, GO term enrichment analysis (Fisher's exact test), KEGG enrichment for DEGs across temperate and tropical cultivars root transcriptome

To predict and analyse the function of the 2239 differentially expressed transcripts, all transcript's sequences were first aligned against those sequences in the NCBI non-redundant (NR) protein database using BLASTx. A total of 1598 (71.37%) significant BLAST top hits with functional annotation were returned with a cut-off E-value of $1e-5$. The same BLAST algorithm ($E\text{-value} < 1e-5$) was used for comparison and to get a similar sequence and corresponding annotation with a recovery of 10 hits per sequence.

Based on Blastx analysis of 2239 DEGs nucleotide query sequences was searched against NCBI's non-redundant NR database through Q Blast. Annotation stringency was decided later in the annotation procedure. Furthermore, we set the hsp filter to 33 to avoid hits where the length of the matching region was smaller than 100 nucleotides. After blast is completed, we visualized different charts (Data distribution, similarity, e-value and species distributions) to get clarity about the quality of the query sequences and the blast procedure. The Blastx statistics is discussed below.

4.11.1 Characteristics of the homology search of common DEGs from root transcriptome between temperate and tropical cultivars against the NR database

a) Data distribution:

Among the total 2239 DEGs, no sequences were left without analysis or without processing. The InterProScan (protein family prediction DB) gave information regarding functional analysis of proteins annotated all 2239 DEGs, classifying these proteins into families and predicting important sites and domains. 641 sequences did not find any hit for these sequences after BLAST step. 1592 sequences with blast had successful best alignment between the query sequence and our sequence, among which,

442 had Blast hits, 89 (1150) with GO mapping, 1061 sequences were detected with GO annotation, whereas, 6 sequences had annotation from InterProScan Fig. 17a and 17b.

b) Sequence Similarity distribution

The similarity distribution chart shows the distribution of all calculated sequence similarities (percentages), and the overall performance of the alignments. The similarity distribution of the top hits in the NR database shows that most sequences have blast similarity values of 50–60% or higher (Fig. 17c). 600 query sequences had 100% similarity for found sequences after blast step. This information is useful for choosing the annotation cut-off parameter at the annotation on step, and suggests that taking a value of 50 would be adequate.

c) Species distribution

Furthermore, the Species distribution chart (Fig. 17d) gives a list of the different species to which most sequences were aligned during the BLAST step. For species distribution, a great majority of sequences were homologous to genes from *Daucus carota sub spp. sativus* sequences (89.07%) followed by the *Daucus carota* (0.70%) and *Vitis vinifera* (0.70%), *Camellia sinensis* (0.62%) and 6.09% of the distinct sequences had the top matches (first hit) trained with sequences from the *Daucus carota sub spp. sativus* (Fig. 17d).

d) Mapping

Mapping is the process of retrieving GO terms associated to the hits obtained after a BLAST search. Blast2GO performs different mapping steps to link all 1,592 BLAST hits to the functional information stored in the Gene Ontology database. GO database contains several million functionally annotated gene products for hundreds of different species. All annotations are associated to an Evidence Code which provides information about the quality of this functional assignment. Note that commonly IEA (electronic annotation) is overwhelmed in the mapping results. However, the contribution of this and other experimental evidences type of annotations are important to the finally assigned annotation of the query set. Out of 1592 blasted hits, 1150 sequences had mapping.

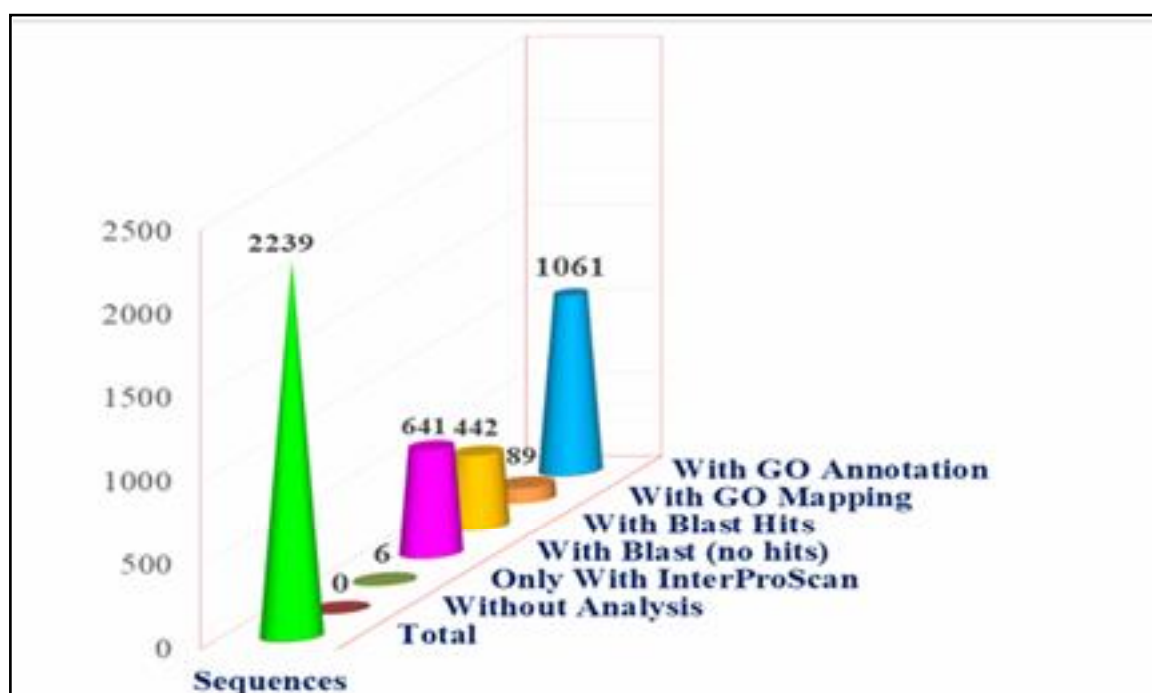


Fig. 17a: Data distribution of un-blasted, blasted, mapped and annotated sequences over the whole data-set of 2239 common DEGs

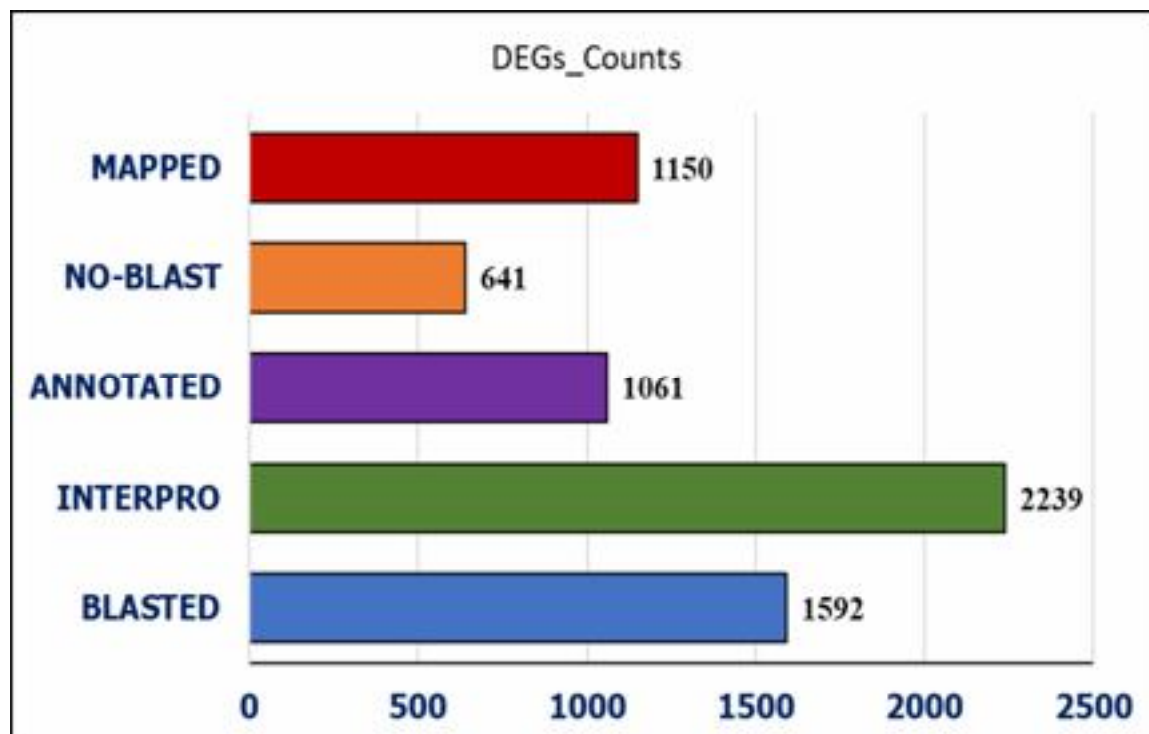


Fig. 17b: Data distribution of un-blasted, blasted, mapped and annotated sequences over the whole data-set of 2239 common DEGs

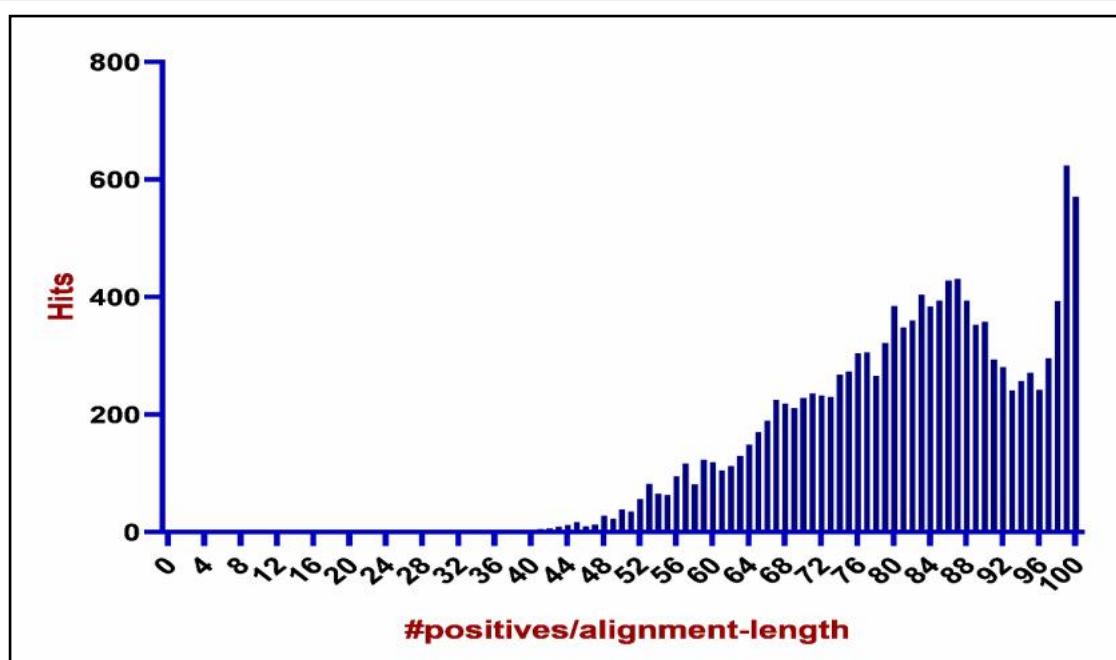


Fig. 17c: Sequence Similarity distribution of the BLAST hits for each common DEGs sequence in the NR database with a cut off E-value of 1.0×10^{-5} for all matching hsp

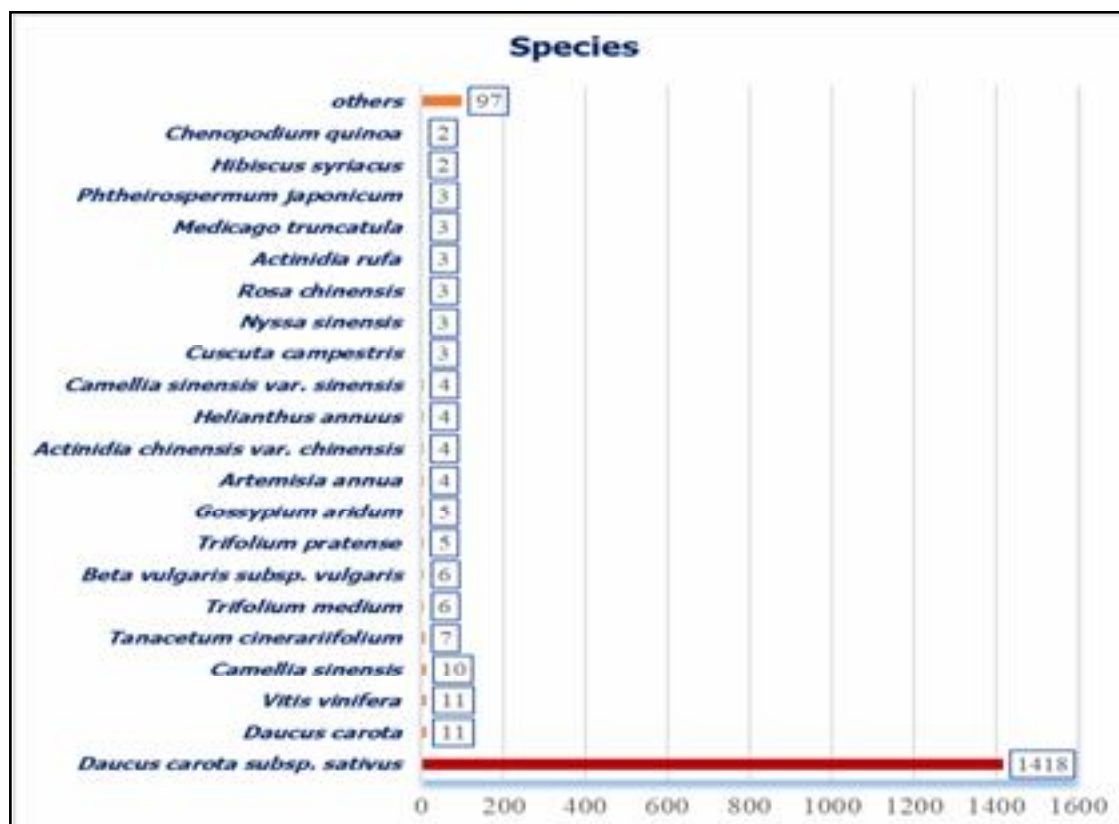


Fig. 17d: Species distribution of the BLAST hits for each Blasted DEGs in the NR database

e) Gene Function Annotation

This is the process of selecting GO terms from the GO pool obtained by the mapping step and assigning them to the query sequences. GO annotation is based on annotation rule (AR) found on ontology terms. The most specific annotations are with certain level of reliability and stringency. Each candidate GO is computed based on annotation score (AS). AS is comprised of two additive terms, first direct term (DT) representing highest hit similarity of the GO weighted by a factor corresponding to its evidence code (EC). The second term (AT) of the annotation score provides the possibility of abstraction where annotation to a parent node when several child nodes are present in the GO candidate collection.

The gene functional annotation following the above rule in Blast2GO annotated total of 1598 (71.37%) sequences with functional annotation among which, 442 had Blast hits, 89 (1150) with GO mapping, 1061 sequences were detected with GO annotation, whereas, 6 sequences had annotation from InterProScan (Fig. 17 a, 17b).

f) GO Functional classification of common DEGs identified from root transcriptome

Gene Ontology (GO) analysis was performed to functionally classify the temperate and tropical carrot cultivar root transcripts, GO terms were assigned to each transcript using Blast2GO based on the best Blastx hit from the NR database. Among the 2239 common DEGs a total of 1598 (71.37%) transcripts were successfully annotated and classified into three GO categories: biological process, cellular component and molecular function, and assigned to total 111 GO functional groups across categories.

Distribution of Gene ontology terms in the assembly of tropical and temperate carrots root transcriptome across BP, CC, MF is presented in Fig. 18. Among the three main GO categories the biological process domain was the most abundant with 62 biological categories (723 sequences), whereas, cellular components showed 16 categories (710 sequences) and molecular function domain with 26 categories (812 sequences) as presented in detail in Fig. 19a, 19b and 19c respectively. Within the biological process domain, “cellular process” (606, 85.81%) and “metabolic process”

(506; 69.98%) were the two most representative subcategories. In the cellular component category, common DEGs related to “cellular anatomical entity” (708; 99.71%) and “intercellular anatomical structure” (407; 57.32%) were most abundant. Meanwhile, in the molecular function category, the majority of DEGs were involved in “binding” (488; 60.09%) and “catalytic activity” (12,326; 58.74%) were the largest proportion represented GO terms Figure 20c. These results suggested that many kinds of enzyme pathways were active in both temperate and tropical carrots.

g) Fisher’s Exact test for GO terms enrichment in up and downregulated DEGs with total DEGs as reference set

Further, GO-enriched categories identified by Fishers Exact Test (FET) with regard to temperate cultivars biological processes (BP) mainly comprised of Photosynthesis as overrepresented category. While, thylakoid, thylakoid membrane, photosynthetic membrane and photosystem were the most enriched cellular components found to be overrepresented (CC). However, Transferase activity, transporter activity, catalytic activity, protein kinase activity, anion binding, small molecule binding were the most enriched under-represented molecular functions (MF) categories identified in temperate cultivars (Fig. 20a).

GO-enriched categories identified by Fishers Exact Test (FET) with regard to tropical cultivars biological processes (BP) mainly included of protein phosphorylation as overrepresented category with 34 genes of test sets and 12 gene sets from reference set. Whereas, ‘intrinsic component of membrane’, integral component of membrane, and ‘cellular anatomical entity’ followed by ‘intrinsic & integral component of the membrane’ enrichment in reference and test sets, were the most enriched cellular components found to be overrepresented (CC). While, ‘Catalytic activity’, ‘organic cyclic compound binding’, heterocyclic compound binding, ‘protein kinase activity’ *etc.* were the most enriched overrepresented molecular functions (MF) categories identified in tropical cultivars. Important to note that all the three enriched domains BP, CC, MF were found to have overrepresented categories in tropical cultivars (Fig. 20b).

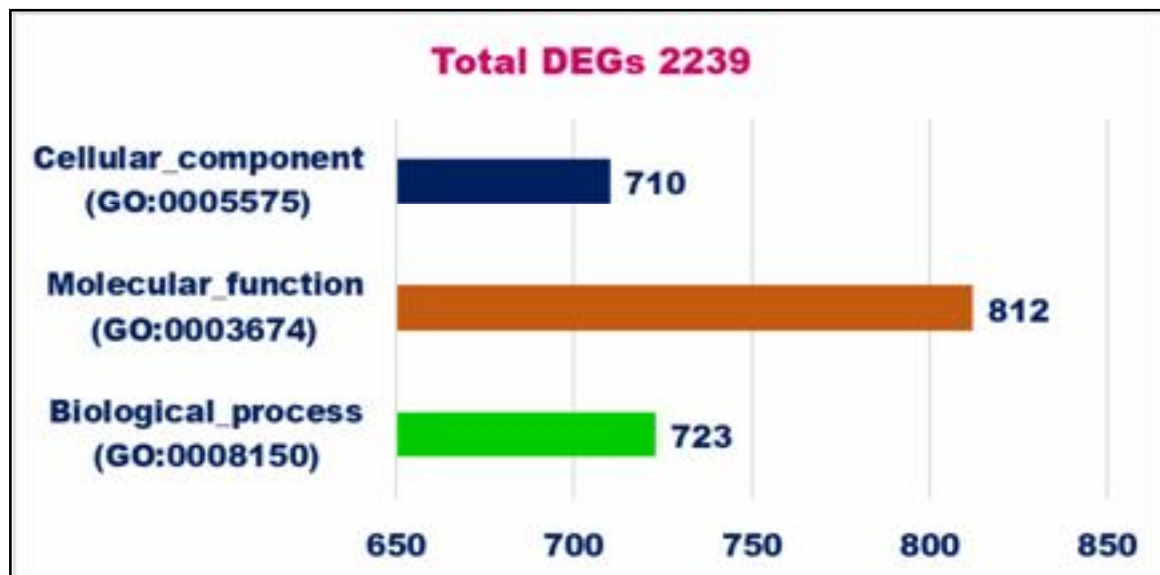


Fig. 18: Distribution of Gene ontology terms in genes exclusively present in the assembly of tropical and temperate carrots root transcriptome (A) Biological process, (B) Cellular component, (C) Molecular function

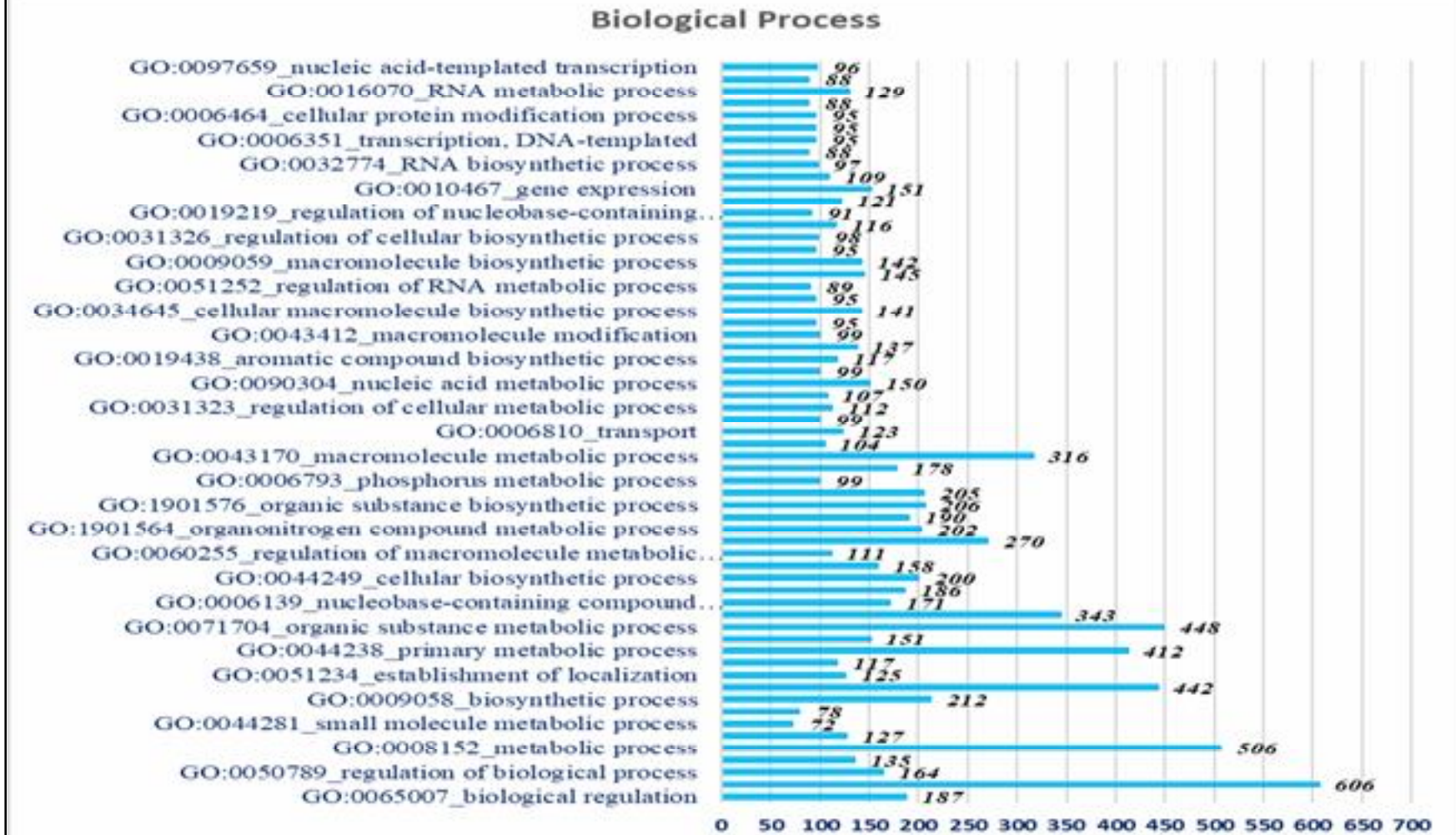


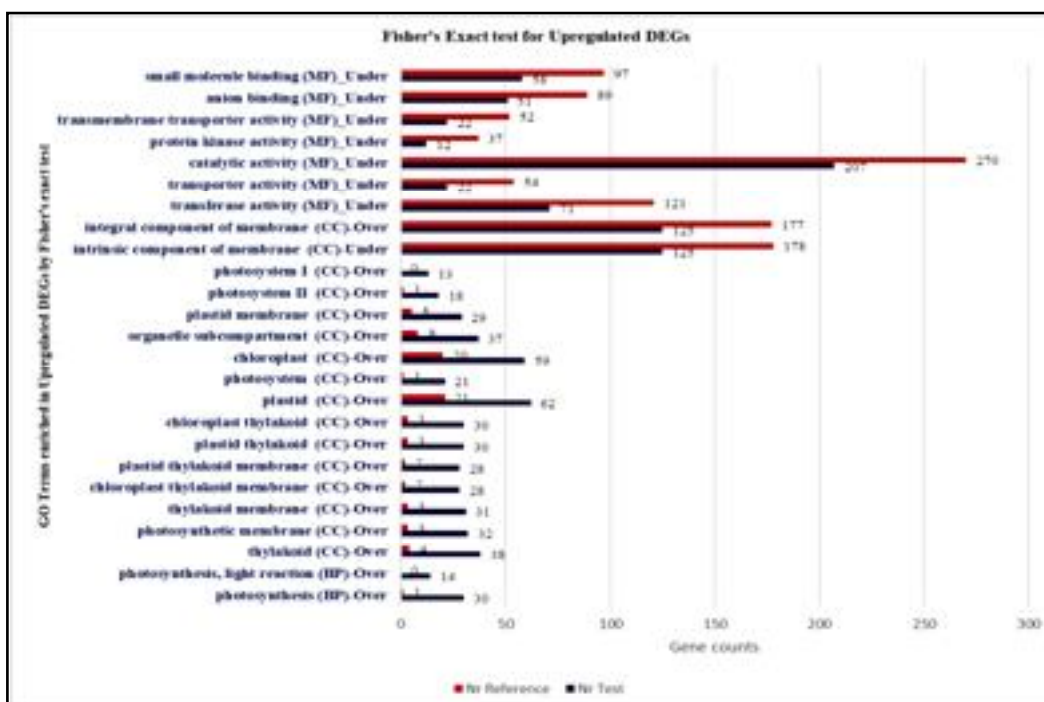
Fig. 19a: GO categorization of DEGs identified across tropical and temperate cultivars for biological process (BP) in root transcriptome based on *De novo* assembly



Fig. 19b: GO categorization of DEGs identified across tropical and temperate cultivars for cellular component (CC) in root transcriptome based on *De novo* assembly

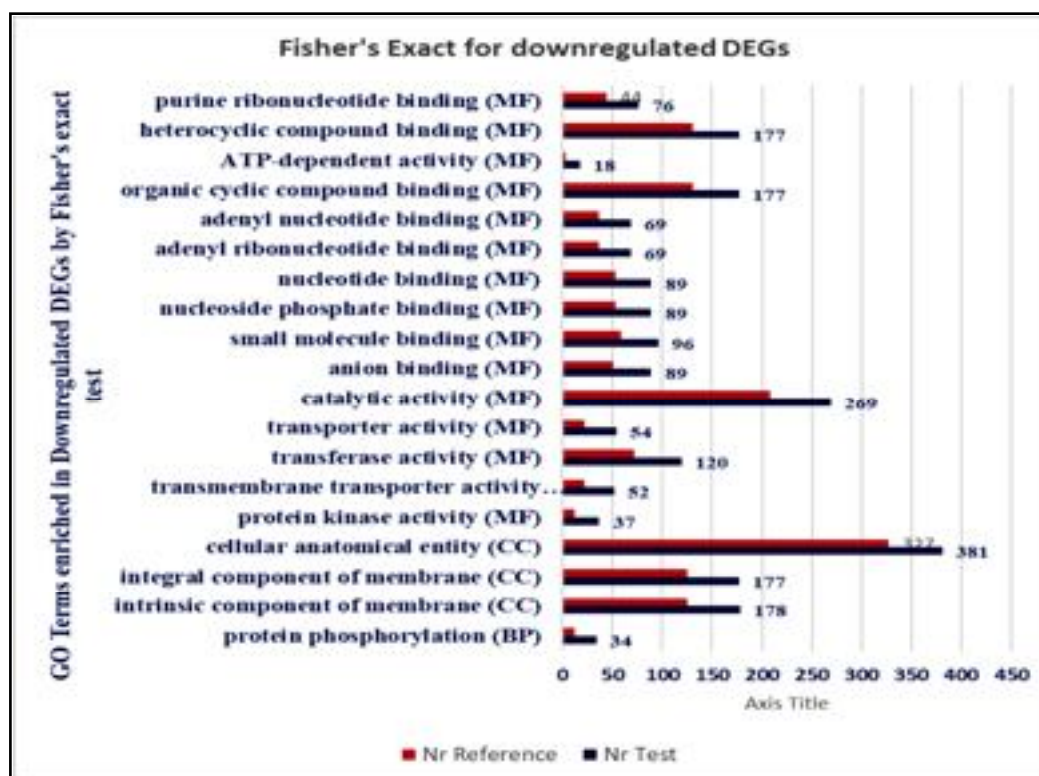


Fig. 19c: GO categorization of DEGs identified across tropical and temperate cultivars for molecular function (MF) in root transcriptome based on *De novo* assembly



*BP and CC terms were overrepresented and MF were under represented in temperate up-regulated cultivars/tropical down regulated

Fig. 20a: Fisher's exact test for upregulated DEGs as test set and total DEGs as reference set for GO groups



*All terms were overrepresented in temperate down-regulated cultivars/tropical upregulated

Fig. 20b: Fisher's exact test for down regulated DEGs as test set and total DEGs as reference set for GO groups

h) GO enrichment and KEGG pathway mapping of common DEGs identified from root transcriptome

The KEGG (Kyoto Encyclopaedia of Genes and Genomes) metabolic system is a group of metabolic maps which represents current understanding of biomolecular interaction networks (Kanehisa *et al.*, 2002; Kanehisa *et al.*, 2004). KEGG pathway analysis with a cut off of $E < 1E-5$ was used to further understand the transcriptome data. Of the 1592 annotated genes (13.96%) with significant matches from the database were assigned 76 KEGG pathways (Table 21) with the top enriched pathways included 'starch and sucrose metabolism', glycolysis/gluconeogenesis, inositol phosphate metabolism, purine metabolism, pentose phosphate pathway, steroids, terpenoid biosynthesis, photosynthesis and metabolism of many of the vitamins (Thiamine, vitamin B6, folate, riboflavin) *etc.* Enzymes that were involved in multiple pathways of secondary metabolites, biosynthetic pathways were identified such as reductase; kinase, different transferases, decarboxylase, hexose diphosphatase *etc.* were top enriched representative pathways (Table 21a). These results indicated that the photosynthesis, carbon metabolism and terpenoid pathways and metabolism of many vitamins were active in carrot root of temperate and tropical carrots, and the corresponding genes would be candidate genes for all above mentioned biosynthetic pathways.

4.11.2 Functional annotation and classification of the DEGs across temperate and tropical carrot FPL tissue transcriptome

To predict and analyse the function of 1671 common DEGs (898 Up and 773 downregulated in temperate and tropical FPL tissue transcriptome respectively) were subjected for Blast2GO to determine the functions associated with 1671 common DEGs and to increase the likelihood of interpreting biological processes.

a) GO categorization

Further, information on biological processes (BP), molecular functions (MF) and cell components (CC) is organized into structured vocabulary, known as gene ontology (GO), that could describe the important pathways between temperate and tropical FPL tissue transcriptome as explained below.

Table 21a: KEGG pathways identified across temperate and tropical cultivars root transcriptome

Sl. No.	Pathways	Pathway ID	#Enzs in Pathway
1	Glycolysis / Gluconeogenesis	map00010	8
2	Glyoxylate and dicarboxylate metabolism	map00630	5
3	Glutathione metabolism	map00480	4
4	Tryptophan metabolism	map00380	2
5	Pyruvate metabolism	map00620	3
6	Cyanoamino acid metabolism	map00460	2
7	Photosynthesis	map00195	2
8	Seleno compound metabolism	map00450	2
9	Fatty acid elongation	map00062	2
10	Styrene degradation	map00643	1
11	Lipoic acid metabolism	map00785	2
12	Galactose metabolism	map00052	4
13	Amino sugar and nucleotide sugar metabolism	map00520	3
14	Biosynthesis of unsaturated fatty acids	map01040	2
15	D-Alanine metabolism	map00473	1
16	Metabolism of xenobiotics by cytochrome P450	map00980	2
17	Aminoacyl-tRNA biosynthesis	map00970	7
18	Tyrosine metabolism	map00350	2
19	Histidine metabolism	map00340	1
20	Oxidative phosphorylation	map00190	4
21	Taurine and hypotaurine metabolism	map00430	2
22	Cysteine and methionine metabolism	map00270	3
23	Folate biosynthesis	map00790	2
24	Biosynthesis of siderophore group nonribosomal peptides	map01053	1
25	Inositol phosphate metabolism	map00562	8
26	PD-L1 expression and PD-1 checkpoint pathway in cancer	map05235	1
27	beta-Alanine metabolism	map00410	1

Sl. No.	Pathways	Pathway ID	#Enzs in Pathway
28	Glycine, serine and threonine metabolism	map00260	1
29	Starch and sucrose metabolism	map00500	10
30	Drug metabolism – other enzymes	map00983	3
31	Alanine, aspartate and glutamate metabolism	map00250	3
32	Fatty acid biosynthesis	map00061	2
33	Porphyrin and chlorophyll metabolism	map00860	6
34	Pyrimidine metabolism	map00240	1
35	Fructose and mannose metabolism	map00051	7
36	Sesquiterpenoid and triterpenoid biosynthesis	map00909	1
37	Arginine and proline metabolism	map00330	2
38	Isoquinoline alkaloid biosynthesis	map00950	1
39	Human immunodeficiency virus 1 infection	map05170	1
40	Phenylpropanoid biosynthesis	map00940	3
41	Indole alkaloid biosynthesis	map00901	1
42	Lysine degradation	map00310	1
43	Sulfur metabolism	map00920	4
44	Pantothenate and CoA biosynthesis	map00770	2
45	Steroid hormone biosynthesis	map00140	2
46	Nicotinate and nicotinamide metabolism	map00760	3
47	Purine metabolism	map00230	8
48	Vitamin B6 metabolism	map00750	1
49	Glycerolipid metabolism	map00561	5
50	Arginine biosynthesis	map00220	2
51	Arachidonic acid metabolism	map00590	1
52	Drug metabolism – cytochrome P450	map00982	3
53	Retinol metabolism	map00830	1
54	Methane metabolism	map00680	4
55	Ether lipid metabolism	map00565	1
56	Nitrogen metabolism	map00910	6
57	C5-Branched dibasic acid metabolism	map00660	1
58	Pentose and glucuronate interconversions	map00040	5

Sl. No.	Pathways	Pathway ID	#Enzs in Pathway
59	Terpenoid backbone biosynthesis	map00900	1
60	Ubiquinone and other terpenoid-quinone biosynthesis	map00130	2
61	Butanoate metabolism	map00650	2
62	Glucosinolate biosynthesis	map00966	1
63	Riboflavin metabolism	map00740	1
64	Glycerophospholipid metabolism	map00564	6
65	Thiamine metabolism	map00730	2
66	Cutin, suberine and wax biosynthesis	map00073	2
67	Carbon fixation pathways in prokaryotes	map00720	3
68	Steroid biosynthesis	map00100	2
69	Carbon fixation in photosynthetic organisms	map00710	5
70	Ascorbate and aldarate metabolism	map00053	3
71	Pentose phosphate pathway	map00030	8
72	Other glycan degradation	map00511	1
73	Peptidoglycan biosynthesis	map00550	1
74	Anthocyanin biosynthesis	map00942	1
75	Propanoate metabolism	map00640	2
76	Citrate cycle (TCA cycle)	map00020	4

The assembled transcripts were annotated using Gene Ontology (GO) and KEGG databases. In this study, 1671 common DEGs between temperate and tropical FPL tissue transcriptome were searched against the GO database. Out of 1671 common DEGs in temperate or downregulated in tropical FPL tissue, 1184 (70.85%) were successfully blasted with minimum e-value 10^{-6} and with 100% Interpro mapping. GO analysis classified into three GO categories: biological process (66), cellular component (15), and molecular function (28) with the corresponding number GO terms (Fig. 21 a,b,c).

As shown in Fig. 21a, GO for BP, assignments which fell under biological process ranked the highest (67), followed by metabolic process (62), and cellular and metabolic process term (54). In this, category, “organic substance metabolic process” (37), “metabolic process” (50), “reproduction” (28) were the few other most representative subcategories. Along with this, other important BP included were cellular metabolic process, stimulus to biotic and abiotic stress, response to stimulus, biosynthetic process *etc* (Fig. 21a).

In the cellular component category, interestingly most of the terms were associated with “cellular component” (75), “cellular anatomical entity” (43), “intracellular anatomical structure” (37), “organelle” (35) *etc* (Fig. 21b). Thee cellular componenet category ith nrched GO terms clearly indicates that most of the DEGs are involved in preparation of plant for floral transition to reproductive from vegetative phase. Among the molecular function, terms like molecular function (65), binding (45) were highly abundant. Transferase activity (11), catalytic activity (22) was other terms enriched in MF (Fig 21c). These results suggested that many kinds of enzyme pathways were active in FPL tissue transcriptome.

b) KEGG annotation of DEGs of FPL transcriptome-*De novo* based

To understand metabolic level enrichment, 1671 DEGs were subjected to KEGG pathway analysis from FPL transcriptome. A total of 40 pathways found for identified DEGs in KEGG database. The top enriched pathways included pyruvate metabolism (4) methane metabolism (3), tyrosine metabolism (3), glycine, serine and threonine metabolism (3), carbon fixation in photosynthetic organisms (3) *etc* (Table 21b).

Table 21b: KEGG pathways identified across temperate and tropical cultivars FPL transcriptome

Sl. No.	Pathway	ID	#Seqs
1	Indole alkaloid biosynthesis	map00901	1
2	Glyoxylate and dicarboxylate metabolism	map00630	1
3	Phenylpropanoid biosynthesis	map00940	2
4	Glycolysis / Gluconeogenesis	map00010	2
5	Glutathione metabolism	map00480	1
6	Tryptophan metabolism	map00380	2
7	Pyruvate metabolism	map00620	4
8	Naphthalene degradation	map00626	1
9	Cyanoamino acid metabolism	map00460	2
10	Betalain biosynthesis	map00965	1
11	alpha-Linolenic acid metabolism	map00592	1
12	Stilbenoid, diarylheptanoid and gingerol biosynthesis	map00945	1
13	Valine, leucine and isoleucine degradation	map00280	1
14	Drug metabolism - cytochrome P450	map00982	2
15	Amino sugar and nucleotide sugar metabolism	map00520	1
16	Retinol metabolism	map00830	1
17	Methane metabolism	map00680	3
18	Metabolism of xenobiotics by cytochrome P450	map00980	2
19	Flavonoid biosynthesis	map00941	1
20	One carbon pool by folate	map00670	1
21	Phenylalanine metabolism	map00360	2
22	Monoterpenoid biosynthesis	map00902	1
23	Biosynthesis of various plant secondary metabolites	map00999	1

24	Tyrosine metabolism	map00350	3
25	Terpenoid backbone biosynthesis	map00900	1
26	Histidine metabolism	map00340	1
27	Tropane, piperidine and pyridine alkaloid biosynthesis	map00960	1
28	Chloroalkane and chloroalkene degradation	map00625	1
29	PD-L1 expression and PD-1 checkpoint pathway in cancer	map05235	1
30	Steroid biosynthesis	map00100	1
31	beta-Alanine metabolism	map00410	1
32	Glycine, serine and threonine metabolism	map00260	3
33	Fatty acid degradation	map00071	1
34	Starch and sucrose metabolism	map00500	2
35	Drug metabolism - other enzymes	map00983	1
36	Carbon fixation in photosynthetic organisms	map00710	3
37	Pentose phosphate pathway	map00030	1
38	Fructose and mannose metabolism	map00051	1
39	Propanoate metabolism	map00640	1
40	Isoquinoline alkaloid biosynthesis	map00950	2

4.12 Transcriptional Factors (TF) identified in root transcriptome in the present study

Transcription factors (TFs) is a group of protein molecule that specifically bind to the 5' end upstream sequence of targeted gene to ensure genes expression under specific temporal and spatial condition. On the other hand, transcription factors are associated with the various developmental stages and play crucial roles in cellular processes (Riechmann and Ratcliffe, 2000). Our *de novo* assembled unigenes and DEGs were subjected to TF identification in the present study as detailed below.

Both Unigenes (44,856) and 2239 common DEGs (1149 and 1090 Up and down regulated respectively) were separately subjected to transcription factors identification by plantTF database. Total of 477 TF from unigenes and total 61 TF from 2239 DEGs consisting of 28 and 33 TF from 1149 up and 1090 down regulated DEGs respectively were identified in the present study (Fig. 22a). From the unigenes, 45 different classes (from 477 TF) of TF were identified with the highest TF included were bHLH, ERF with 40 sequences each, WRKY and C2H2 TF with 28 sequences each, MYB-related and bZIP with 27 each, GRAS (23), NAC (20) *etc* were most common TF (Figure23b). Similarly, TFs distributed throughout whole root transcriptome of tropical and temperate carrots were identified and summarised to understand the main implications of this information for the breeding and development of marker-assisted selection strategies, with particular focus on characteristics of direct agronomical interest by comparing with the available bibliography (Bianchi *et al.*, 2015). Details of identified TFs in the current study in root transcriptome is mentioned in Table 22; Fig. 22b.

In case of DEGs upregulated in temperate cultivar, 19 TF classes (28 total TF) were identified wherein, the top TF included were AP2, GATA, GRF were higher in number compared to tropical. Among the tropical cultivar upregulated DEGs also 19 different TF were identified wherein, bHLH, ERF, NAC and WRKY transcription factors were higher compared to temperate. Whereas, TF that were common in both the cultivars were bZIP (2 TF each), HSF (2 TF), MYB related (one), HD-ZIP (one). The identification TFs in carrot presents important agronomical and molecular advantages, as the availability of complete genome data in carrot (comparatively small 480Mbp) make it more interesting for identifying candidate TFs linked various agronomic traits and promoter identification from genomic data.

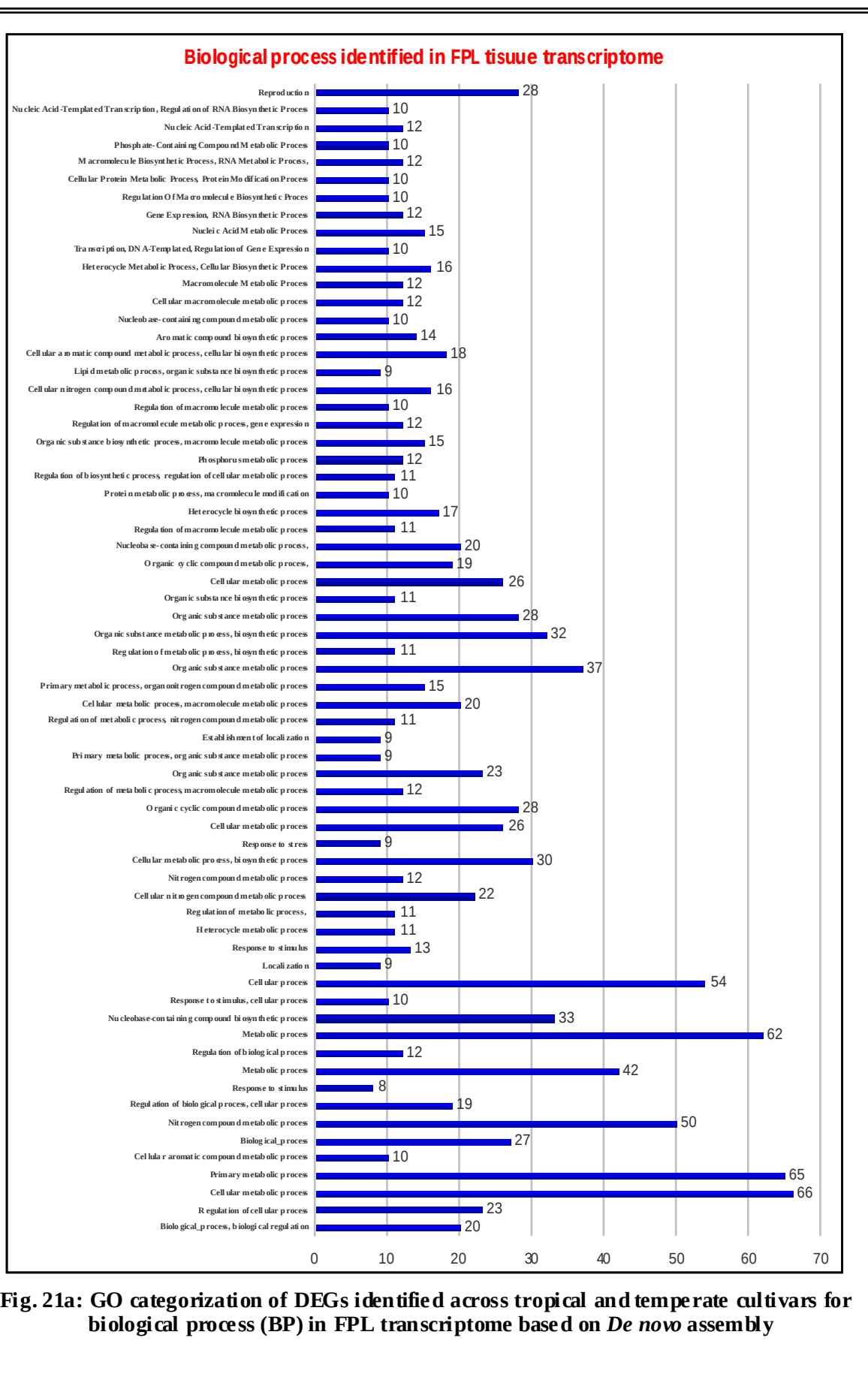


Fig. 21a: GO categorization of DEGs identified across tropical and temperate cultivars for biological process (BP) in FPL transcriptome based on *De novo* assembly

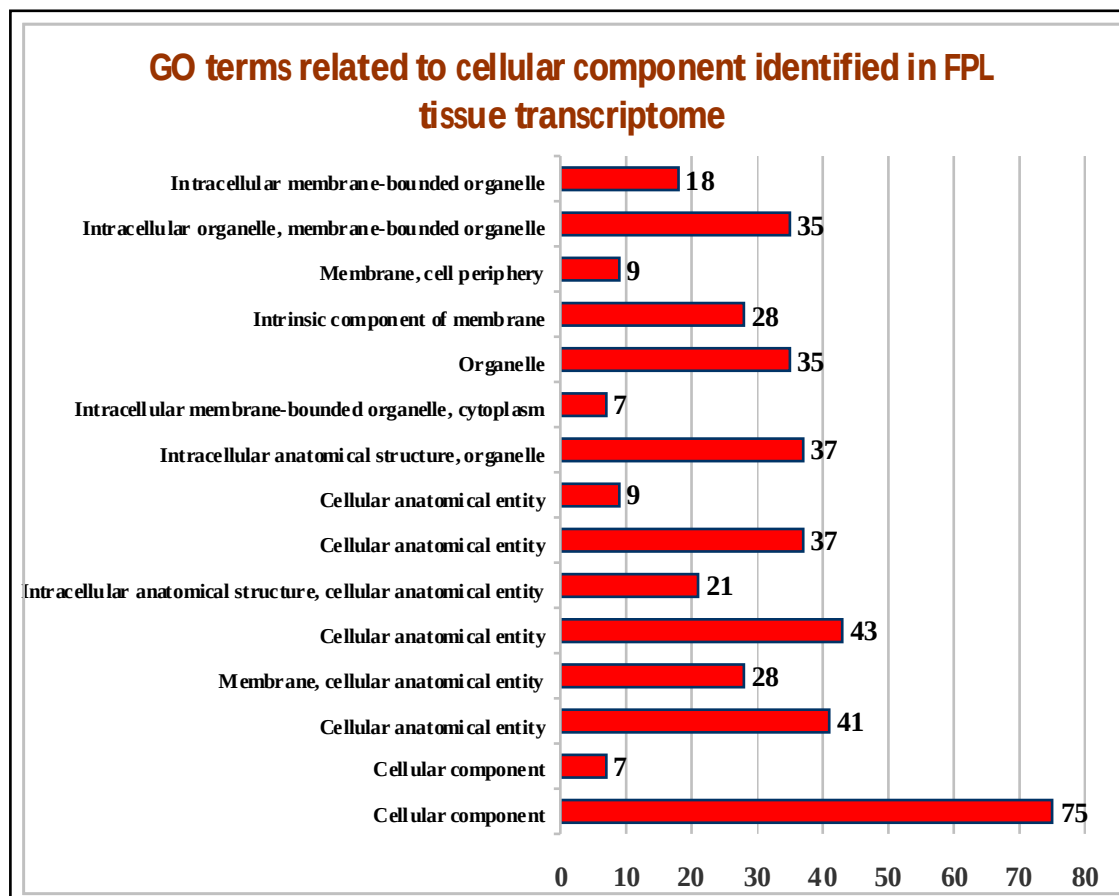


Fig. 21b: GO categorization of DEGs identified across tropical and temperate cultivars for cellular component (CC) in FPL transcriptome based on *De novo* assembly

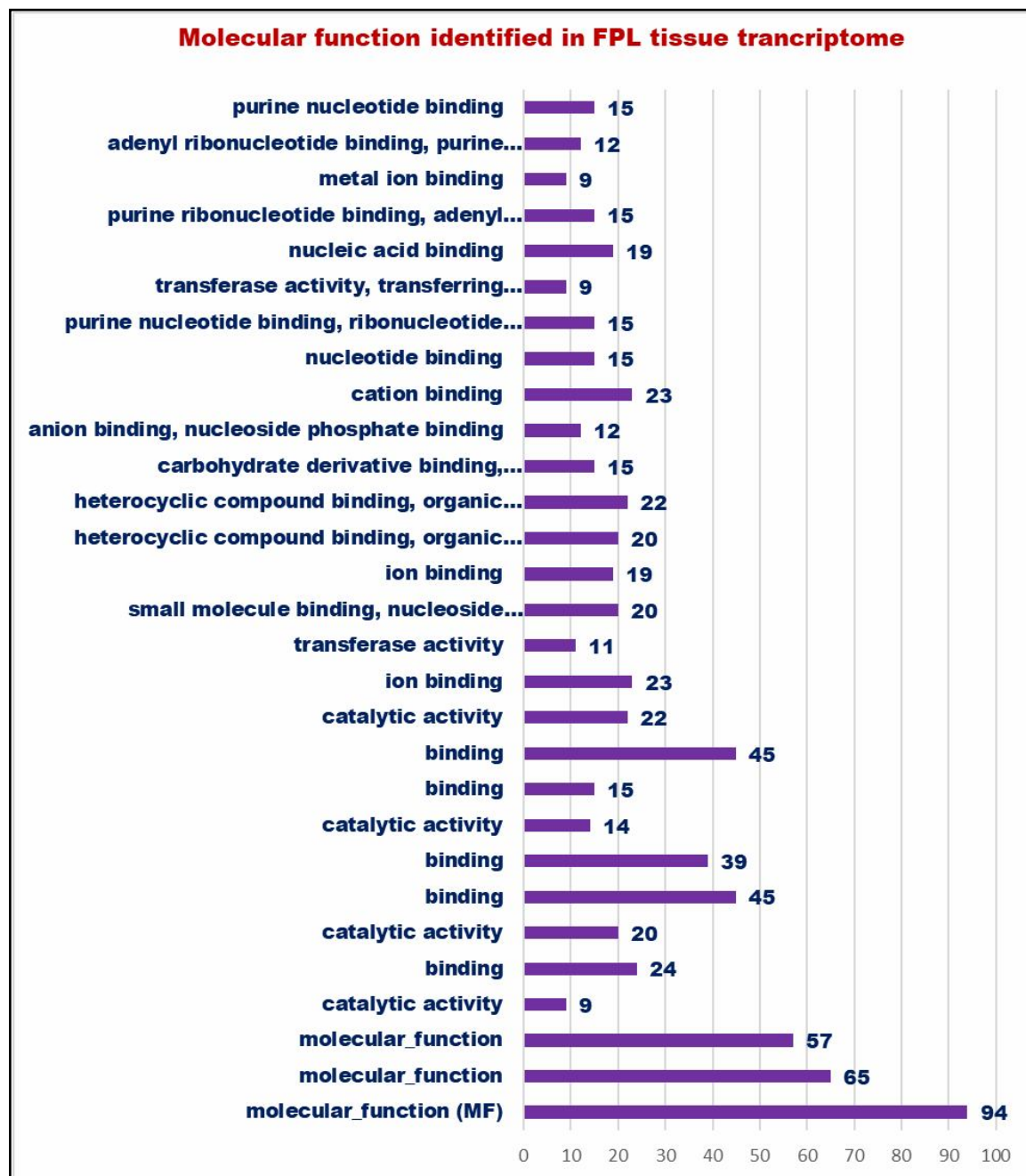


Fig. 21c: GO categorization of DEGs identified across tropical and temperate cultivars for molecular function (MF) in FPL transcriptome based on *De novo* assembly

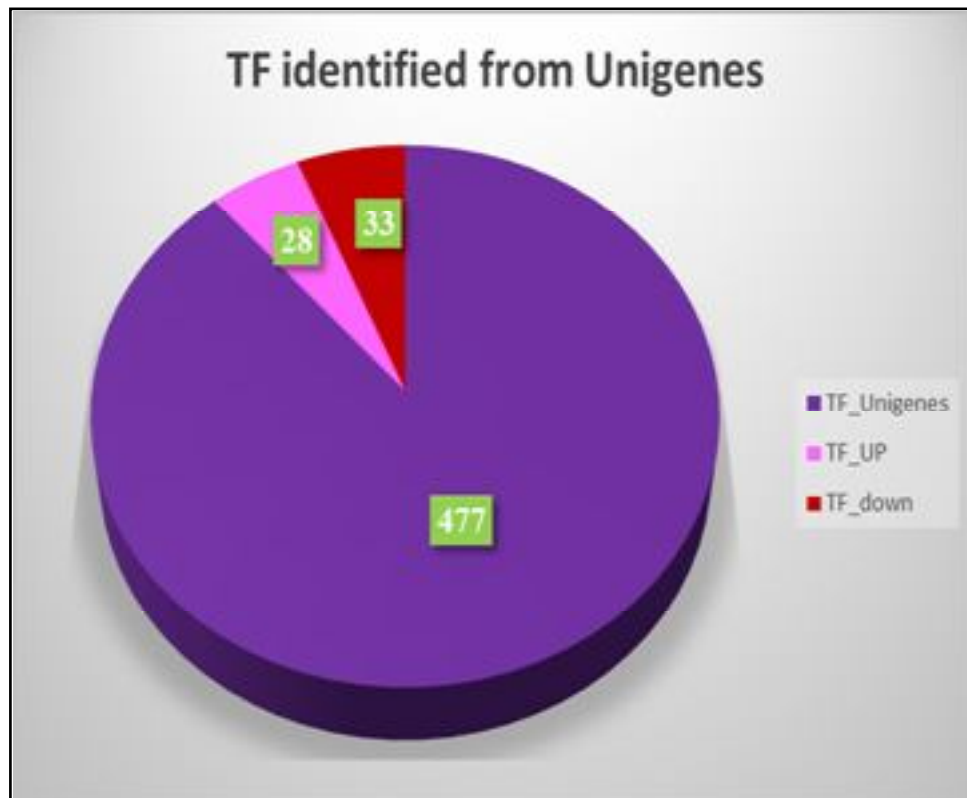


Fig. 22a: Total transcriptional factors identified for the unigenes (44856), up and down regulated DEGs among root transcriptome

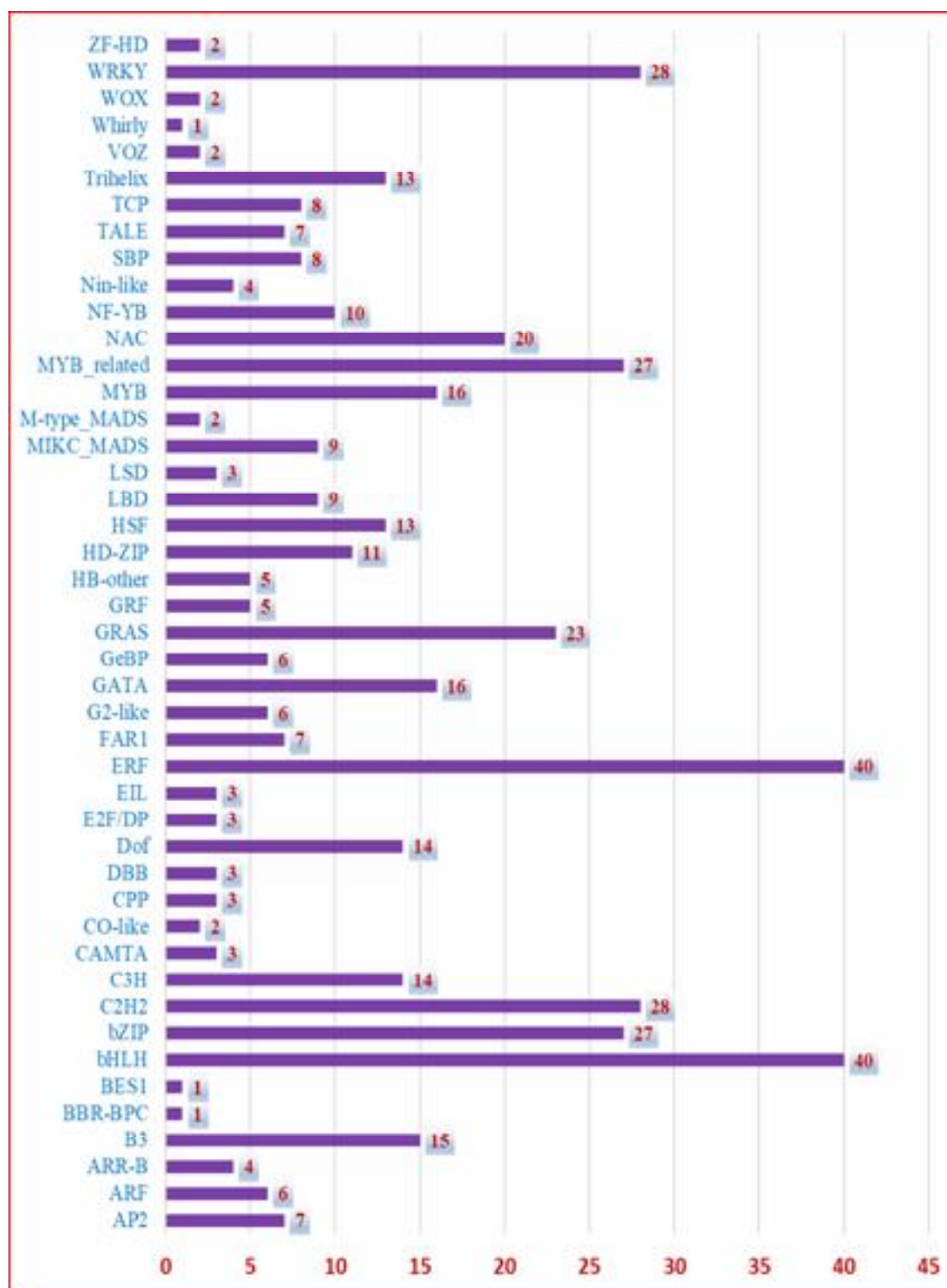


Fig. 22b: Transcriptional factors identified for root transcriptome from unigenes lists

Table 22: Number and classes of transcriptional factors identified by unigenes and 2239 common DEGs in root tissues

TF	Unigenes	TF Up regulated in temperate	TF Up regulated in tropical
AP2	7	3	1
ARF	6		
ARR-B	4		
B3	15		1
BBR-BPC	1		
BES1	1		
bHLH	40	1	4
bZIP	27	2	2
C2H2	28	1	1
C3H	14		
CAMTA	3		
CO-like	2		
CPP	3		
DBB	3	1	1
Dof	14	1	1
E2F/DP	3		
EIL	3		
ERF	40	3	4
FAR1	7		
G2-like	6		1
GATA	16	2	
GeBP	6		
GRAS	23		
GRF	5	2	
HB-other	5		
HD-ZIP	11	1	1
HSF	13	2	2
LBD	9		1
LSD	3	1	
MIKC_MADS	9		3
M-type_MADS	2	1	1
MYB	16	1	
MYB_related	27	1	1
NAC	20	1	3
NF-YB	10		
Nin-like	4	1	1
SBP	8	1	1
TALE	7		
TCP	8		
Trihelix	13		
VOZ	2		
Whirly	1		
WOX	2		
WRKY	28	2	3
ZF-HD	2		

4.12.1 Transcriptional Factors (TF) identified in FPL tissue transcriptome in the present study

Unigenes (49,978) and 1671 common DEGs (898 Up and 773 downregulated respectively) were separately subjected to transcription factors identification by plantTF database. We found total of 539 sequences annotated to TF in FPL tissue transcriptome, and these have been classified into 53 TF families (Fig. 23) among unigenes. TF families expressed in the FPL tissue transcriptome were high compared to the matured storage root transcriptome of the carrot. Among common 1671 DEGs, 29 loci encoded for TF with 15 different TF families. Of these 29 TFs differentially expressed, 8 TF were induced upregulated in temperate or downregulated in tropical cultivar. While, 21 TF were downregulated in temperate, in other words upregulated in tropical cultivars. This suggests the possible major transcriptional reprogramming occurred to manifest these two cultivars phenotype and flowering pathway (Fig. 23).

Of all the TF families in FPL tissue unigenes, bHLH (44) and ERF (37) and C₂H₂ (34) TF family were highly enriched. There were 34 genes that code for C₂H₂ TFs were induced in FPL tissue transcriptome followed by MYB (28), HD-ZIP (21) bZIP (29), C₃H (21), MYB_related (21), NAC (21), WRKY (18), GATA (16), Trihelix (16), TALE (14) were other major TF identified. Among other TF, AP2 (8), GRAS (21), B3(14), G2-like (12), Dof (11), TCP (8), E2F/DP (7), MIKC_MADS (11), M-type_MADS (4), LFY (1) and SBP (9) were also identified in unigenes sequences.

4.12.2 Distribution of differentially expressed transcription factor families among 1671 common DEGs found in FPL tissue transcriptome

Among the TF identified for DEGs, many of them (21 out of 29) were upregulated in tropical or downregulated in temperate. The temperate downregulated DEGs were as follows. YABBY (2), MIKC-MADS (5), HD-ZIP (3), MYB-Related (5), MYB (1), bZIP (1), TALE (1), B3(1) and NF-YA (1), TCP (1) were downregulated in temperate. Among the upregulated TF in temperate cultivars, bZIP (3), WRKY (1), ERF (1), C₂H₂ (2) and MYB (1) were found as upregulated. Details of the identified TFs among the DEGs in the current study are mentioned in Table 23 and among unigenes are presented in Figure 23.

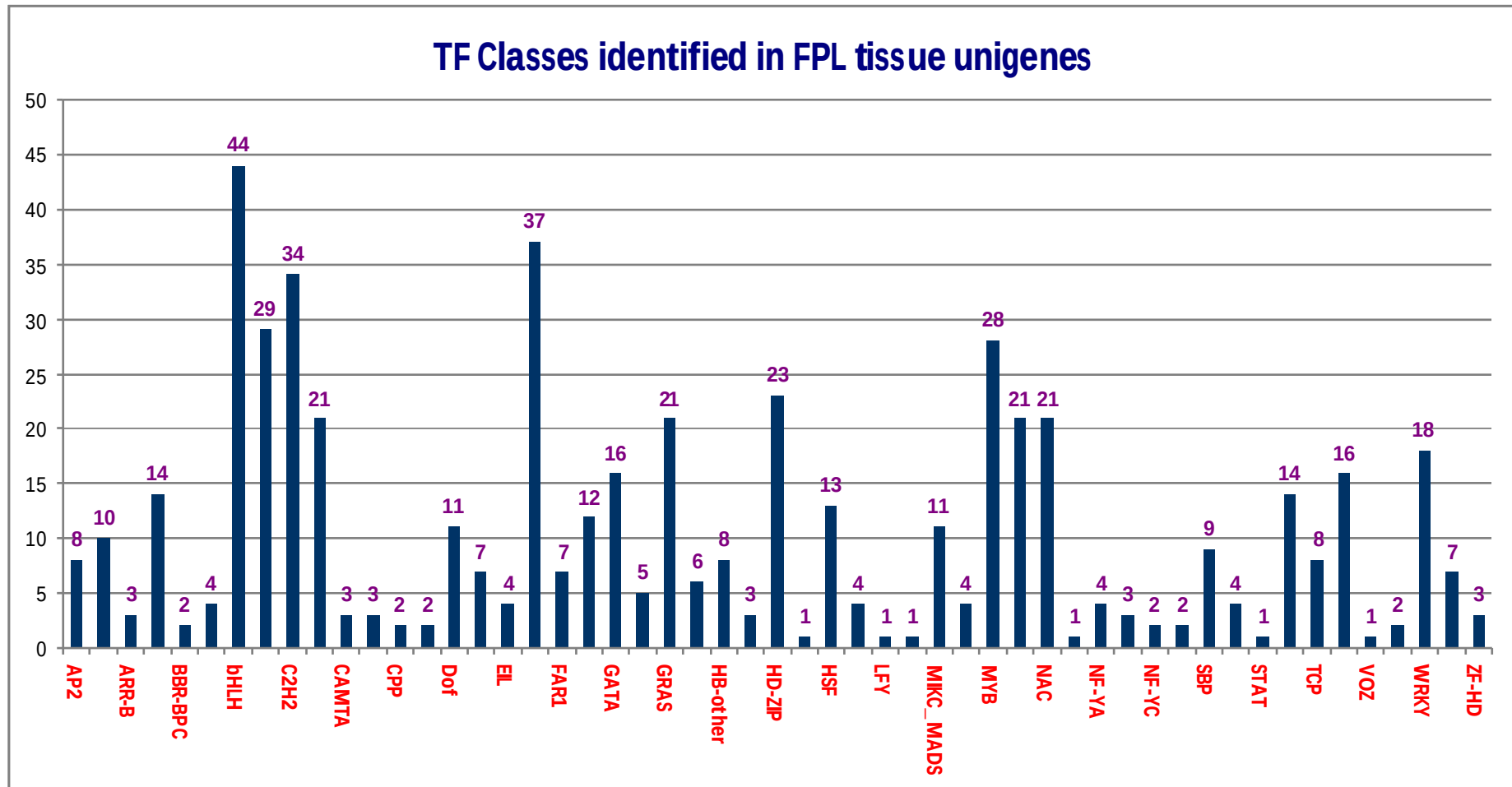


Fig. 23: Transcriptional factors identified for FPL tissue transcriptome from unigenes lists

Table 23: List of differentially expressed TF families identified among temperate and tropical FPL tissue DEGs

S No	Gene ID	TF_family	Direction of DEG
1	TRINITY_DN2048_c0_g1_i3	bZIP	Up
2	TRINITY_DN23226_c0_g1_i9	bZIP	Down
3	TRINITY_DN30810_c0_g1_i1	bZIP	Up
4	TRINITY_DN38879_c0_g1_i1	NF-YA	Down
5	TRINITY_DN3953_c0_g2_i3	WRKY	Up
6	TRINITY_DN4335_c1_g1_i2	YABBY	Down
7	TRINITY_DN4627_c0_g1_i2	bZIP	Up
8	TRINITY_DN5412_c0_g1_i22	TCP	Down
9	TRINITY_DN8131_c0_g1_i1	YABBY	Down
10	TRINITY_DN39427_c0_g1_i1	B3	Down
11	TRINITY_DN33945_c1_g2_i1	ERF	Up
12	TRINITY_DN13219_c0_g1_i2	MIKC_MADS	Down
13	TRINITY_DN5301_c0_g1_i10	MIKC_MADS	Down
14	TRINITY_DN646_c0_g1_i8	MIKC_MADS	Down
15	TRINITY_DN646_c0_g2_i2	MIKC_MADS	Down
16	TRINITY_DN646_c1_g3_i3	MIKC_MADS	Down
17	TRINITY_DN15412_c0_g3_i1	TALE	Down
18	TRINITY_DN2803_c0_g1_i19	HD-ZIP	Down
19	TRINITY_DN28833_c0_g1_i1	HD-ZIP	Down
20	TRINITY_DN4219_c0_g1_i2	HD-ZIP	Down
21	TRINITY_DN2687_c1_g1_i1	C2H2	Up
22	TRINITY_DN5070_c0_g1_i2	C2H2	Up
23	TRINITY_DN10080_c0_g2_i2	MYB	Up
24	TRINITY_DN2275_c0_g1_i2	MYB	Down
25	TRINITY_DN1087_c0_g1_i10	MYB_related	Down
26	TRINITY_DN2683_c0_g1_i1	MYB_related	Down
27	TRINITY_DN37612_c0_g1_i1	MYB_related	Down
28	TRINITY_DN5870_c0_g1_i1	MYB_related	Down
29	TRINITY_DN7219_c0_g1_i1	MYB_related	Down

Floral homeotic transcription factors (TFs) act in a combinatorial manner to specify the organ identities in the flower. However, the architecture and the function of the gene regulatory network (GRN) controlling floral organ specification is still poorly understood. In particular, the interconnections of homeotic TFs, microRNAs (miRNAs) and other factors controlling organ initiation and growth have not been studied systematically so far. Here, using a combination of TF binding from FPL tissue transcriptome data, we reconstructed the dynamic GRN controlling floral meristem development and organ differentiation in carrot.

4.13 Genetic variant analysis

4.13.1 Microsatellite and SNP Distribution Analysis in storage root transcriptome of tropical and temperate cultivars

The high-throughput transcriptome sequencing provides excellent resources toward the discovery of cost-effective and polymorphic genetic markers (SSRs, SNPs, Indel). We identified total 4,437 SSR markers of which 520 were compound SSRs by using MISA tool for 44856 unigenes in both tropical and temperate carrot root transcriptome. The average density of perfect microsatellites was found to be one SSR per 1.85kb. Based on the motif repetition, these microsatellites were further categorized into mononucleotide (1239), followed by di-(2016), tri- (873), tetra-(180), Penta-(56) and hexanucleotides motifs (73) ((Table 24). The percentage of perfect SSR types identified between storage root transcriptome of temperate and tropical cultivars is presented in Fig. 24a. (A/C)_n motif was the dominant for mononucleotide, (AG/AC/AT)_n for di- while (AAG/ATC/ACG)_n for trinucleotide microsatellites. Among tetranucleotide repeats, ATAC was the most abundant followed by AAAT Fig. 24b.

4.13.2 SNP discovery between storage root tissue transcriptome of temperate and tropical carrots

SNPs endure the ability to produces high-density genetic maps, association mapping and molecular markers with the promise of lower cost and error rate. In this study, putative variants were called by aligning the raw reads with the non-redundant

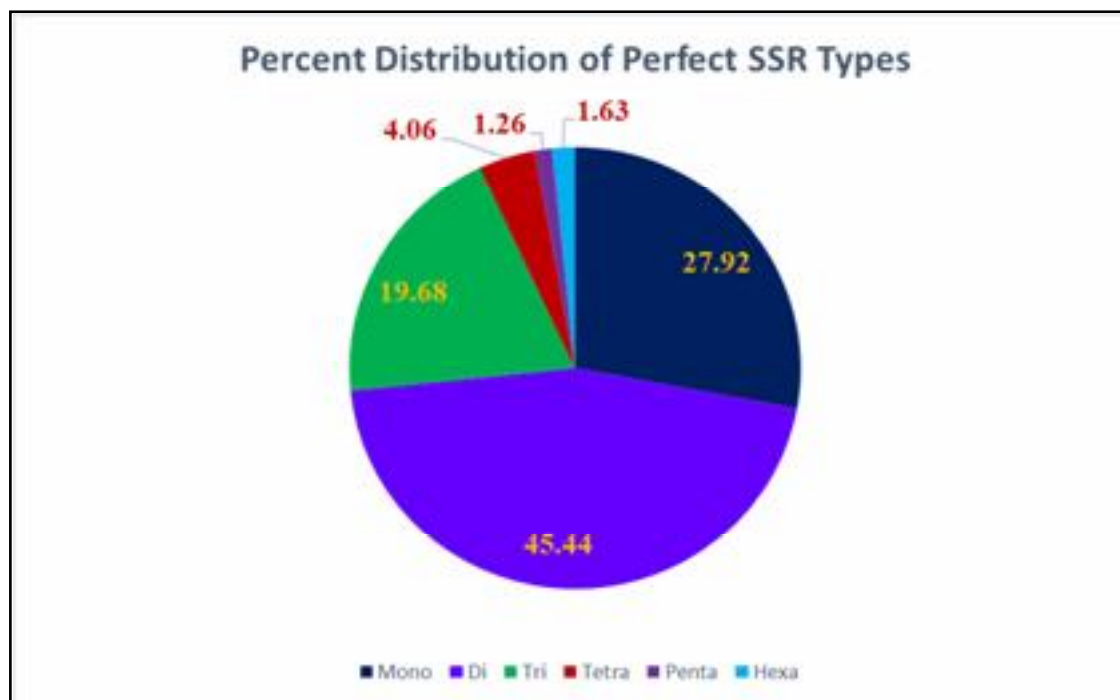


Fig. 24a: Percentage of different perfect SSRs distribution in root transcriptome

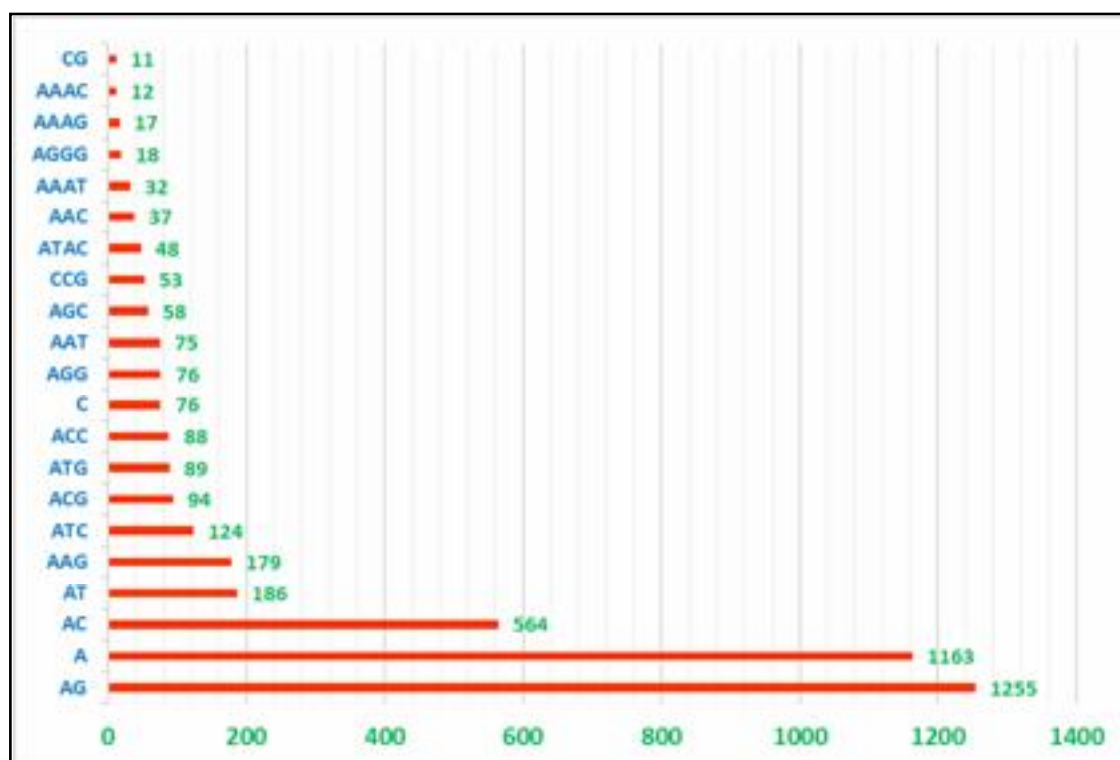


Fig. 24b: Microsatellite distribution analysis and top abundant perfect SSRs found in root transcriptome

Table 24: The summary of perfect microsatellite types identified in root tissue transcriptome

Type	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
Mono	1239	20735	27.92	16.74	28.18	471.66
Di	2016	39696	45.44	19.69	45.86	902.96
Tri	873	14964	19.68	17.14	19.86	340.39
Tetra	180	3108	4.06	17.27	4.09	70.7
Penta	56	1215	1.26	21.7	1.27	27.64
Hexa	73	2028	1.65	27.78	1.66	46.13
Total	4437	81746	100.01	120.32	100.92	1859.48

de novo assembled reference database. A total of 236071 SNPs were identified of which 134335 (56.90 %) transitions (Ts) and 101736 (43.10 %) transversions (Tv) and 1.32 Ts/Tv ratio was identified for root transcriptome data in the present study from six paired -end RNA-seq libraries. Among the transitions, T>C and A>G (~17.5 %) are more frequent than C>T and G>A conversions (10.87 %). Where as in transversions, C>A and G>T (~6 %) was more common than other type of transversions. Transitions and transversion frequencies including variations are listed in Fig. 25 and Table 25.

Total of 20866 indels were identified among the tropical and temperate carrot root transcriptome, single base deletions (-1) were 45% (9505) followed by single base addition (1bp) with 40.45 % (8435). Only 4.99% of the indels were belonging to addition of two bases (2bp) Fig. 25.

4.13.3 Genic microsatellite identification in FPL transcriptome of temperate and tropical cultivars

In order to analyse the microsatellites distribution in the FPL transcriptome, the 49,978 unigenes generated from FPL tissue were used to mine potential genic microsatellites that were defined as mono- to hexanucleotide motifs with a minimum of three repetitions. A total of 7877 potential genic simple sequence repeats (SSRs) were identified in 49,978 unigenes. The average density of perfect microsatellites was found to be one SSR per 2479.88bp/Mb. For the motif types, the di-nucleotide type (6365) was the most abundant followed by tri (750).

Based on the motif repetition, these microsatellites were further categorized into mononucleotide (196), followed by di-(6365), tri- (750), tetra-(152), Penta-(33) and hexanucleotides motifs (381) (Table 27). AG dinucleotide motif was the dominant for dinucleotide repeats followed by AA. while AAG was most abundant for trinucleotide microsatellites. Among tetranucleotide repeats, ATAC was the most abundant followed by AAAT presented in Fig. 26.

4.13.4 SNP discovery between FPL tissue transcriptome of temperate and tropical carrots

In this study, putative variants were called by aligning the raw reads with the non-redundant *de novo* assembled FPL reference database. A total of 445144 SNPs

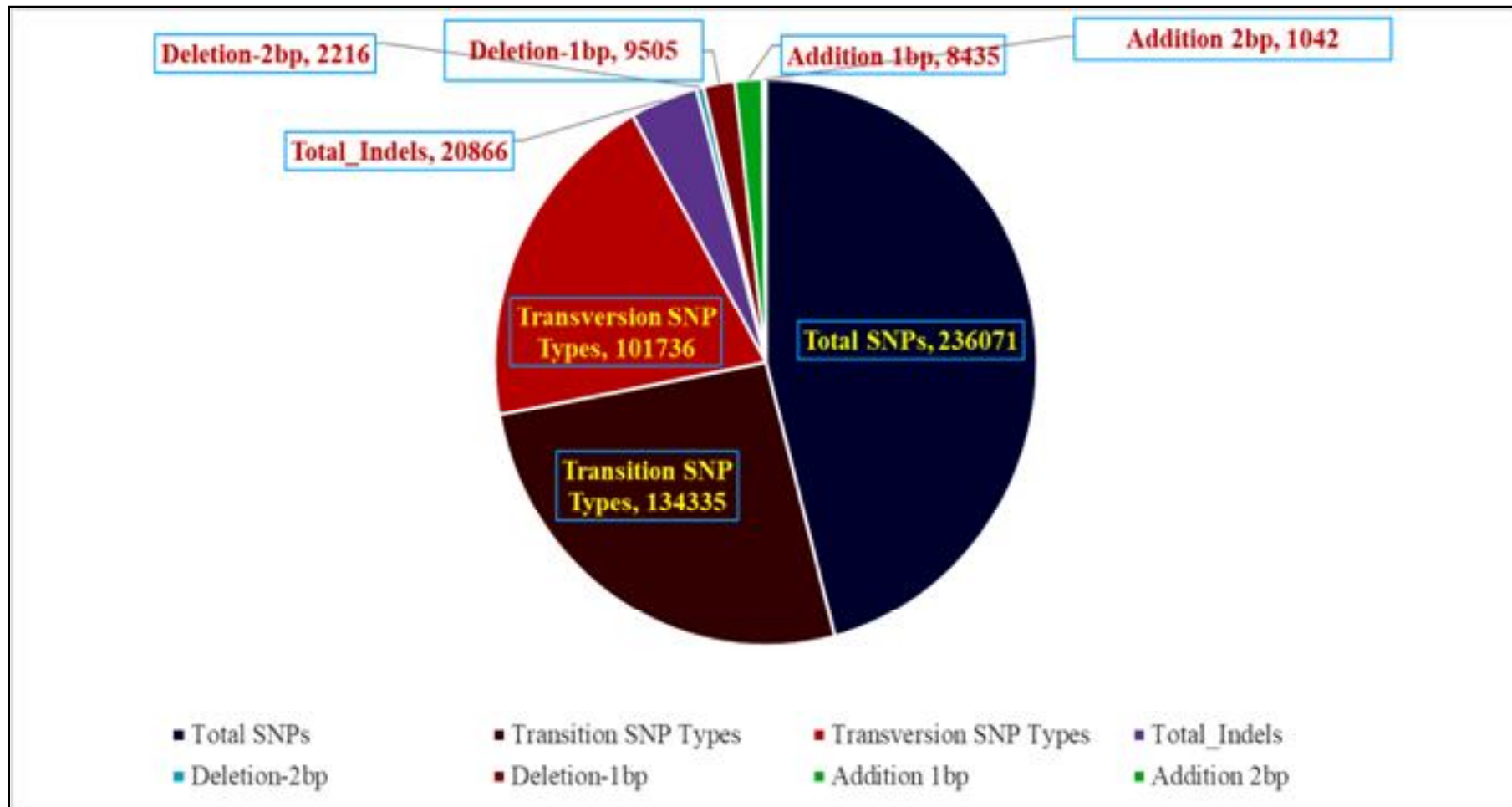


Fig. 25: Identification of SNPs (Transition, Transversion) and Indels (insertions and deletions) in root transcriptome data

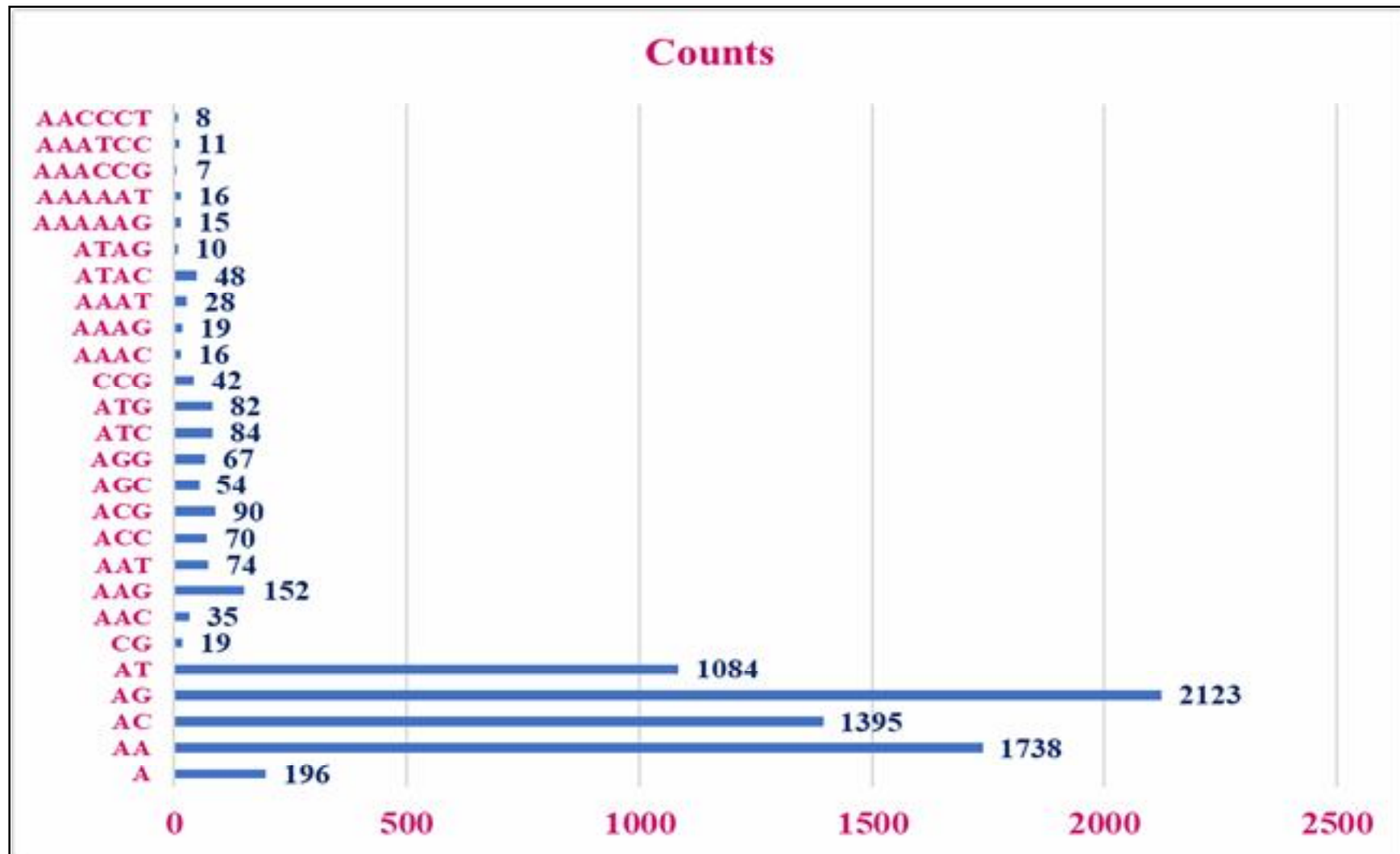


Fig. 26: Microsatellite distribution analysis and top abundant perfect SSRs found in FPL transcriptome

Table 25: List of SNPs and Indels identified in root transcriptome across temperate and tropical

Parameters	Temp_R1	Temp_R2	Temp_R3	Trop-R1	Trop-R2	Trop-R3	Common
Number of samples:	1	1	1	1	1	1	6
number of records:	96663792	108130186	104467826	108255969	102768321	108795922	148053230
number of no-ALTs:	96153217	107445474	103814242	107468486	102077779	108000342	145514149
number of SNPs:	495643	664993	635131	767368	672475	774714	2464153
number of indels:	14932	19719	18453	20115	18067	20866	74928
number of multiallelic sites:	495835	665323	635382	767710	672729	775023	2466541
number of multiallelic SNP sites:	495643	664993	635131	767368	672475	774714	2464153
Transitions (ts)	308758	418021	392976	477334	418458	481784	1534378
Transversion (tv)	215228	281230	280272	337964	289124	335025	1182288
ts/tv ratio	1.43	1.49	1.4	1.41	1.45	1.44	1.3
A>C	27907	42219	37865	46483	38433	44121	163771
A>G	91214	129248	117016	146280	125551	144095	479007
A>T	27878	38406	36484	41457	36758	42500	149348
C>A	31874	30801	37117	48908	41789	48640	161676
C>G	20839	29600	29514	33086	28156	32707	117889
C>T	63648	79943	79851	92302	83635	96742	289186
G>A	64258	80764	80561	93148	84585	98407	290359
G>C	20144	29332	28314	32103	27324	31823	115963
G>T	31164	30341	36443	48194	41137	48216	159489
T>A	27602	38250	36721	41303	36862	42531	149183
T>C	89638	128066	115548	145604	124687	142540	475826
T>G	27820	42281	37814	46430	38665	44487	164969
Indel Distribution							
(Deletion)-2	1717	2201	2087	2232	2035	2216	9445
(Deletion)-1	6699	8914	8256	9146	8161	9505	36207
(Addition)1	5993	7993	7475	8110	7255	8435	28315
(Addition)2	733	975	904	993	887	1042	3607

Table 26: The summary of perfect microsatellite types identified in FPL tissue transcriptome

Type	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
Mono	196	3263	2.49	16.65	4.66	77.56
Di	6365	77666	80.8	12.2	151.3	1846.13
Tri	750	12747	9.52	17	17.83	303.00
Tetra	152	2720	1.93	17.89	3.61	64.65
Penta	33	690	0.42	20.91	0.78	16.40
Hexa	381	7242	4.84	19.01	9.06	172.14
Total	7877	104328	100	103.66	187.24	2479.88

were identified of which 245786 (55.21%) transitions (Ts) and 199358 (44.78%) transversions (Tv) and 1.32 Ts/Tv ratio was identified for FPL transcriptome data in the present study from six paired -end RNA-seq libraries Fig. 27a. Among the transitions, A>G and T>C were more frequent than C>T and G>A conversions and *etc.* Where as in transversions, G>A and C>T was more common than other type of transversions. Transitions and transversion frequencies including variations are presented in Fig. 27b. Total of 25200 indels were identified among the tropical and temperate carrot FPL transcriptome. Fig. 27a.

4.14 Gene Regulatory Network (GRN) analysis for storage root development

Vascular cambium (VC) is the master regulator in deciding the xylem and phloem thickening during secondary storage root formation. Reference genes involved in storage root formation and VC development of *Arabidopsis*, we performed GRN analysis to understand the genes responsible for storage root formation and their interaction with stress responsive genes in temperate and tropical cultivars carrot separately.

A) GRN for storage root development and Photosynthesis regulation in temperate cultivar

In case of temperate carrot cultivar, as discussed earlier, photosynthetic cascades and photosynthesis related genes are upregulated in GO in BP or CC subcategories. GRN constructed using vascular cambium responsive genes in temperate carrots also clustered genes which are interconnected each other are mostly involved in photosynthetic related activities (ex: LSH10, PSAH, PSAN, PSAD, LHCB, PSB, *etc.*), with few stresses' responsive genes like WRKY70, ERF-71, MKK9, MYB15, CAMBP25 *etc.* and with root vascular developmental genes (TRN2, LSH10, ATG13610, AT1G56150, SP1L2, HB6, AT1G66830) Fig. 28.

B) GRN for storage root and stress response in tropical cultivar

GRN analysis in tropical carrot cultivar, in contrast, showed highest number of stress responsive genes/TF (LOX2, LOX3, LOX4, WRKY46, MYB15, ERF4, PUB24,

KIN3, NAC036, NLA *etc*) along with few roots growth-related genes (PERK10, AT2G26380 Leucine-rich repeat (LRR), LSH6, HCA2 *etc*) than photosynthesis network genes (0) or carbohydrate metabolism (NANA) as presented in Fig. 28b.

4.15 Identification of candidate genes for root adaptability and flowering related traits

The objective of the study was to identify candidate genes involved in root adaptability in terms of root development, vascular cambium patterning, biotic and abiotic stress, photosynthetic ability, carbohydrate metabolism as well flowering genes from temperate and tropical adapted cultivars. The identified genes for the above traits are presented below.

A) Candidate genes identified for root development:

Secondary storage root development and root system architecture is important. A strong and well-established root system plays an important role in plant performance and yield in case of carrot plants. Differentially expressed genes involved in root morphogenesis have been identified between tropical and temperate plants and listed in Table 27a and 27b.

In tropical plants 16 GO terms were assigned for different root growth related genes. Among which 2 gene counts for root development and root morphogenesis, xylem and phloem pattern formation. Root hair elongation, Lateral root formation had two gene counts each in tropical cultivar. Shoot apical meristem development, developmental growth involved in morphogenesis, cotyledon vascular tissue pattern formation and stem vascular tissue pattern formation had one gene count each as shown in Table 27a.

Whereas, for temperate carrots only 5 GO terms were assigned against root epidermal cell differentiation (1), lateral root morphogenesis (2) and lateral root development (2) as presented in Table 27b.

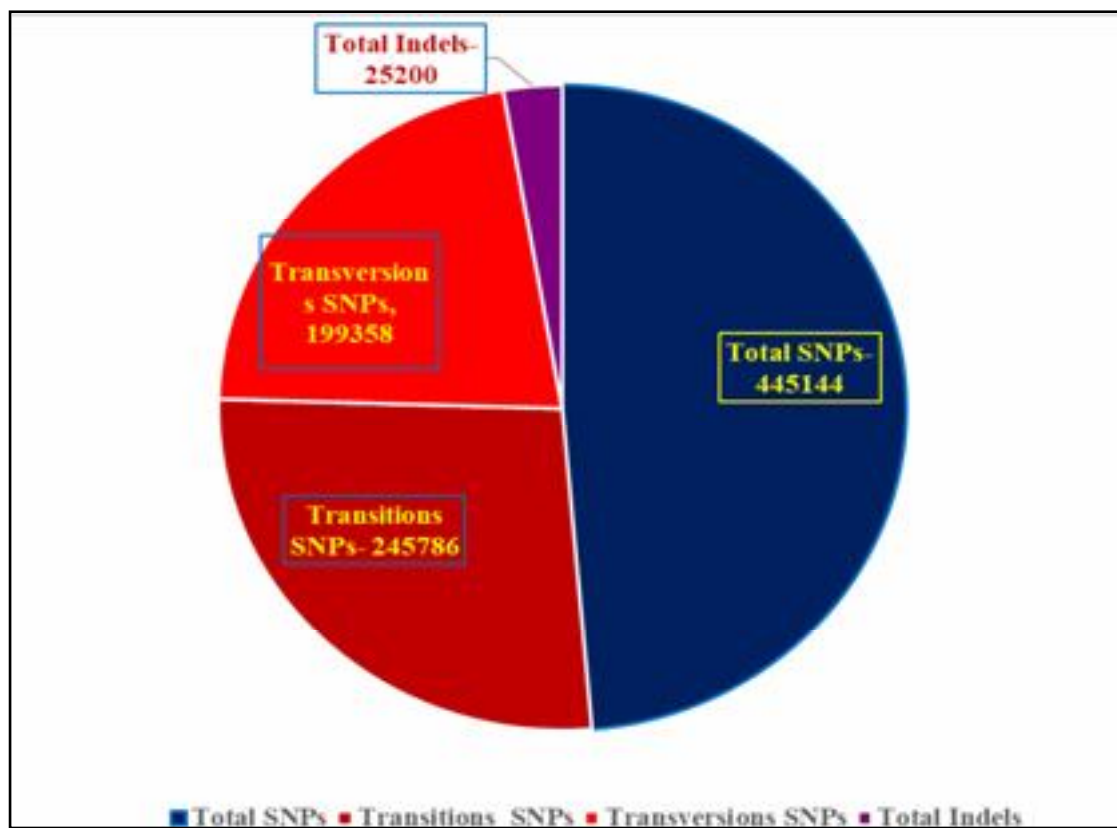


Fig. 27a: Identification of SNPs and Indels in FPL transcriptome

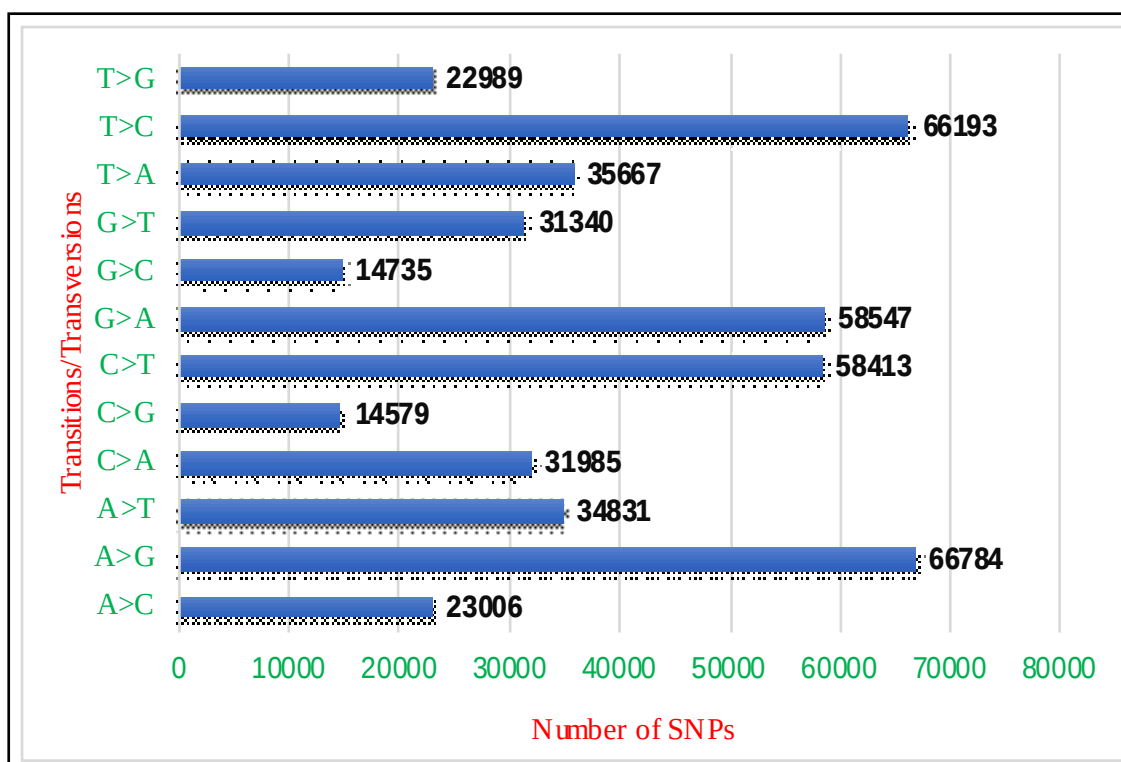


Fig. 27b: Types of SNPs identified in FPL transcriptome

Fig. 28b: Gene regulatory network showing the network of genes filtered for secondary storage root development genes in tropical carrot based on orthology search with *A. thaliana* and network drawn in STRING v. 11.0

Table 27a: Differentially expressed candidate genes identified for ‘root development’ in tropical carrot cultivar

GO_Term	Function	No. of genes/ sequences
GO:1902182	Shoot apical meristem development	1
GO:0060560	Developmental growth involved in morphogenesis	1
GO:0010015	Root morphogenesis	2
GO:0048364	Root development	3
GO:0010089	Xylem development	1
GO:0010051	Xylem and phloem pattern formation	2
GO:0010588	Cotyledon vascular tissue pattern formation	1
GO:0010222	Stem vascular tissue pattern formation	1
GO:0048767	Root hair elongation	2
GO:0010311	Lateral root formation	2

Table 27b: Differentially expressed candidate genes identified for ‘root development’ in temperate carrot cultivar

Go_term	Function	No. of genes/sequences
GO:0010053	Root epidermal cell differentiation	1
GO:0010102	Lateral root morphogenesis	2
GO:0048527	Lateral root development	2

B) Candidate genes identified for carbohydrate and lipid metabolism:

Carrot root carbohydrates, composed mainly of reducing and non-reducing free sugars (mainly glucose, fructose and sucrose) and other more complex carbohydrates have been studied since nearly a century. Sugar content strongly influences carrot flavor. 11 GO terms for candidate genes involved in carbohydrate metabolism were identified in tropical carrots (Table 28a). Polyacetylenes represent a large group of non-volatile lipid compounds produced primarily by members of the *Apiaceae* family. Which have revealed numerous health promoting properties for these compounds. Carrot transcriptomic sequence resources for various key genes and enzymes involved in the biosynthesis of carrot polyacetylenes have been listed below in Table. Nine genes were identified for various synthesis of various polyacetylenes in tropical carrots (Table 28a) whereas, three in temperate carrots (Table 28b)

C) Candidate genes involved in photosynthesis:

Photosynthesis and the transport and accumulation of carbohydrates are the three essential physiological processes of yield formation for any economically important crop. The photosynthetic related candidate genes and relevant GO terms for the same is listed below. Tropical carrots had only two genes for photosynthesis regulation (Table 29a), whereas, 15 different candidate genes were assigned for temperate carrots (Table 29b), indicating the photosynthetic machinery that is well developed in temperate carrots compare to tropical carrots.

D) Candidate genes for constitutive defense response for biotic and abiotic factors

Tropical carrots were found to have a greater number of genes identified for various biotic and abiotic stress responses. A more detailed information on GO terms for target genes to various biotic and abiotic stresses enriched between tropical and temperate carrots storage root in the present study are listed in the Table 30a, 30b.

A total of 41 gene counts expressing constitutionally in tropical carrots were identified for different classes of pathogens. Fungus- 14, bacteria- 10, virus-02,

Table 28a: Differentially expressed candidate genes identified for carbohydrate and lipid metabolism in tropical carrot cultivar

GO_Term	Function	No. of genes/ sequences
Carbohydrate metabolism		
GO:0005983	Starch catabolic process	1
GO:0030244	Cellulose biosynthetic process	1
GO:0006006	Glucose metabolic process	1
GO:0006096	Glycolytic process	1
GO:0005975	Carbohydrate metabolic process	1
GO:0051070	Galactomannan biosynthetic process	1
GO:0010417	Glucuronoxylan biosynthetic process	2
GO:0010439	Regulation of glucosinolate biosynthetic process	2
GO:0010392	Galacto glucomannan metabolic process	1
Lipid metabolism		
GO:0043693	Monoterpene biosynthetic process	1
GO:0051762	Sesquiterpene biosynthetic process	1
GO:0016117	Carotenoid biosynthetic process	1
GO:0016126	Sterol biosynthetic process	1
GO:0016123	Xanthophyll biosynthetic process	1
GO:1905038	Regulation of membrane lipid metabolic process	1
GO:0006636	Unsaturated fatty acid biosynthetic process	1
GO:0005543	Phospholipid binding	1
GO:0036092	Phosphatidylinositol-3-phosphate biosynthetic process	1

Table 28b: Differentially expressed candidate genes identified for carbohydrate and lipid metabolism in temperate carrot cultivar

GO_Term	Function	No. of genes/sequences
GO:0006633	Fatty acid biosynthetic process	1
GO:0016117	Carotenoid biosynthetic process	1
GO:0016123	Xanthophyll biosynthetic process	1

Table 29a: Differentially expressed candidate genes identified for photosynthesis in tropical carrot cultivar

GO_Term	Function	No. of genes/sequences
GO:0015979	Photosynthesis	1
GO:0010109	Regulation of photosynthesis	1

Table 29b: Differentially expressed candidate genes identified for photosynthesis in temperate carrot cultivar

GO_Term	Function	No. of genes/sequences
GO:0015979	Photosynthesis	4
GO:0048564	Photosystem I assembly	1
GO:0009767	Photosynthetic electron transport chain	1
GO:0009768	Photosynthesis, light harvesting in photosystem I	2
GO:0010207	Photosystem II assembly	2
GO:0009769	Photosynthesis, light harvesting in photosystem II	2
GO:0042549	Photosystem II stabilization	1
GO:0010109	Regulation of photosynthesis	1
GO:0010206	Photosystem II repair	1

Table 30a: Differentially expressed candidate genes identified for biotic stress responses in tropical carrot cultivar

GO_Term	Function	No. of genes/sequences
GO:0009626	Plant-type hypersensitive response	1
GO:0009627	Systemic acquired resistance	3
GO:0009816	Defence response to bacterium , incompatible interaction	1
GO:0009817	Defence response to fungus , incompatible interaction	1
GO:0009615	Response to virus	1
GO:0051607	Defence response to virus	1
GO:0045087	Innate immune response	2
GO:0002213	Defence response to insect	1
GO:0042742	Defence response to bacterium	5
GO:0050829	Defence response to Gram-negative bacterium	3
GO:0002229	Defence response to oomycetes	1
GO:0050832	Defence response to fungus	10
GO:0009624	Response to nematode	4
GO:0009617	Response to bacterium	1
GO:0002238	Response to molecule of fungal origin	2
GO:0010421	Hydrogen peroxide-mediated programmed cell death	1
GO:0052544	Defence response by callose deposition in cell wall	1
GO:0010200	Response to chitin	1
GO:0031348	Negative regulation of Defence response	1

Table 30b: Differentially expressed candidate genes identified for biotic stress responses in temperate carrot cultivar

GO_Term	Function	No. of genes/sequences
GO:0042742	Defense response to bacterium	04
GO:0009697	Salicylic acid biosynthetic process	01
GO:0009620	Response to fungus	01
GO:0009753	Response to Jasmonic acid	02

insect- 01, other innate defence mechanism -10 (Table 30a). Whereas, temperate carrot storage root was poorly enriched with different biotic stress responsive classes (four), bacterium (4), fungus (1) Salicylic acid biosynthetic process (01), Response to Jasmonic acid (02). This shows the importance of tropical carrots in biotic stress breeding in carrot (Table 30b).

A total of 72 genes against 21 different classes of abiotic stress were identified in tropical carrots (Table 31a) whereas, 31 genes against 13 different classes of abiotic stress were identified in temperate carrots which are almost half as compared to tropical carrots (Table 31b).

E) Flowering related DEGs identified from FPL tissue transcriptome.

When the list of DEGs functionally annotated was searched for flowering or reproduction or flower morphogenesis terms across the temperate and tropical cultivars, the following results were obtained. A set of target genes identified for floral transition, pollen development, overall flower morphogenesis and seed development were highest in tropical carrots (46), in which 21 were involved in flower regulation and remaining 25 genes were involved seed development as presented in Table 32a. However, in temperate carrots (3) that indicates low enrichment of flower related genes in temperate cultivar. The results show the diverse flowering response and pathways between two cultivars.

4.16 Validation of DEGs identified storage root by RNA-Seq results with quantitative real-time PCR (qRT-PCR)

RNA samples isolated for the RNA sequencing were also used to perform the qRT-PCR analysis to validate the results of RNA-Seq. Nine DEGs were selected from root tissue and five DEGs from FPL tissue for the validation of RNAs-seq identified DEGs by qRT-PCR. FPKM value was calculated for all samples, and the Log₂ ratio of FPKM was used to reflect the fold changes compared with Log₂FC estimated based on CT value of qPCR. These results are shown in Fig. 29a for root tissues RNA-seq DEGs validation, whereas, Fig. 29b shows comparative expression of temperate and tropical cultivars by qPCR. Relative expression changes based on qRT-PCR results were largely found to be in accordance to the RNA-Seq data, although with a few exceptions. Taken together, RNA-Seq results were confirmed by the qRT-PCR analysis.

Table 31a: Differentially expressed candidate genes identified for abiotic stress responses in tropical carrot cultivar

GO_Term	Function	No. of genes/ Sequences
GO:0016036	Cellular response to phosphate starvation	2
GO:0050826	Response to freezing	2
GO:1900036	Positive regulation of cellular response to heat	1
GO:0009651	Response to salt stress	9
GO:0042538	Hyperosmotic salinity response	1
GO:0010193	Response to ozone	1
GO:0009269	Response to desiccation	1
GO:0009631	Cold acclimation	2
GO:0070370	Cellular heat acclimation	1
GO:0006979	Response to oxidative stress	4
GO:0042542	Response to hydrogen peroxide	1
GO:0010421	Hydrogen peroxide-mediated programmed cell death	1
GO:0006970	Response to osmotic stress	3
GO:0009611	Response to wounding	5
GO:0009414	Response to water deprivation	11
GO:0031347	Regulation of defense response	1
GO:0009408	Response to heat	4
GO:1990641	Response to iron ion starvation	1
GO:0045087	Innate immune response	2
GO:0034605	Cellular response to heat	1
GO:0046686	Response to cadmium ion	6
GO:0009409	Response to cold	12

Table 31b: Differentially expressed candidate genes identified for abiotic stress responses in temperate carrot cultivar

GO_Term	Function	No. of genes/ sequences
GO:0009269	Response to desiccation	1
GO:0009408	Response to heat	1
GO:0071731	Response to nitric oxide	1
GO:0006979	Response to oxidative stress	1
GO:0009409	Response to cold	7
GO:0009651	Response to salt stress	3
GO:0009414	Response to water deprivation	1
GO:0009635	Response to herbicide	2
GO:0046686	Response to cadmium ion	2
GO:0030912	Response to deep water	1
GO:0010150	Leaf senescence	2
GO:0046686	Response to cadmium ion	2
GO:0009644	Response to high light intensity	3
GO:0009416	Response to light stimulus	6

Table 32a: Candidate genes involved in flowering and reproduction identified from tropical carrot FPL transcriptome

GO_Term	Function	No. of genes/sequences
GO:0010077	Maintenance of inflorescence meristem identity	1
GO:0010582	Floral meristem determinacy	2
GO:0048510	Regulation of timing of transition from vegetative to reproductive phase	2
GO:0010050	Vegetative phase change	2
GO:0009911	Positive regulation of flower development	1
GO:0080086	Stamen filament development	1
GO:0009555	Pollen development	2
GO:0009860	Pollen tube growth	3
GO:0009846	Pollen germination	2
GO:0010152	Pollen maturation	1
GO:0010483	Pollen tube reception	1
GO:0048481	Plant ovule development	1
GO:0048439	Flower morphogenesis	2
Seed development		
GO:0090378	Seed trichome elongation	3
GO:0009836	Fruit ripening, climacteric	1
GO:0009960	Endosperm development	1
GO:0009793	Embryo development ending in seed dormancy	3
GO:0010154	Fruit development	2
GO:0048316	Seed development	3
GO:0010214	Seed coat development	1
GO:0010262	Somatic embryogenesis	1
GO:0010214	Seed coat development	1
GO:0009793	Embryo development ending in seed dormancy	3
GO:0010262	Somatic embryogenesis	1
GO:0010231	Maintenance of seed dormancy	1
GO:0009845	Seed germination	3
GO:0010187	Negative regulation of seed germination	1

Table 32b: Candidate genes involved in flowering and reproduction identified from temperate carrot FPL transcriptome

GO_Term	Function	No. of genes/sequences
GO:0048481	Plant ovule development	1
GO:0048316	Seed development	1
GO:0090378	Seed trichome elongation	1

In particular, a subset of nine genes (fold change >2) were also evaluated with qPCR. According to current annotation, these genes G2 (ubiquitin-conjugating enzyme E2 20-like; LOC108208449), G3 (polyubiquitin; LOC108227972), G4 (probable WRKY transcription factor53; LOC108209169) are involved in root development. Then, G5 (probable glycosyltransferase At5g03795; LOC108193533) involved in carbon metabolism and G6 (cysteine proteinase inhibitor B; LOC108211283) is involved in seed germination. Whereas, G7 (ATPase WRNIP1; LOC108207021), G8 (DNA repair protein XRCC2 homolog; LOC108216678), G9 (histone deacetylase 6; LOC108200546) are involved in innate immune response (Plant pathogen interaction). While, G1 (mechanosensitive ion channel protein 6-like; LOC108226635) function in homeostasis and signalling pathways, their interactions with other genes related to growth and development, responsive to light, hormone, and stress.

The fold change values from RNA-seq and qRT-PCR for eight of the nine genes (G2 to G9) genes exhibited concordant tendency of gene expression. Eight genes were in agreement with the direction of the change fold. G1 had discordant tendency of expression between the methods of RNA-seq and qPCR expression in opposing directions. As eight genes are concordant and one gene is non-concordant, obtaining concordant results for a higher number of random selections of genes with good correlation between results obtained with qPCR and with RNA-seq (Fig. 29a). This shows that RNA-seq and qRT-PCR shared the similar expression patterns. RNA-seq results were confirmed by the qRT-PCR analysis.

Further, comparison of selected genes expression pattern between temperate and tropical carrots by qPCR showed that ubiquitin-conjugating enzyme E2 20-like and DNA repair protein XRCC2 homolog were the only DEGs found to be up-regulated in temperate carrots. While, remaining genes like probable WRKY transcription factor53 are involved in root development, probable glycosyltransferase involved in carbon metabolism, ATPase WRNIP1 involved in innate immune response (Plant pathogen interaction), mechanosensitive ion channel protein 6-like; LOC108226635) function in homeostasis and signalling pathways are found to be up-regulated in tropical carrots (Fig. 29b).

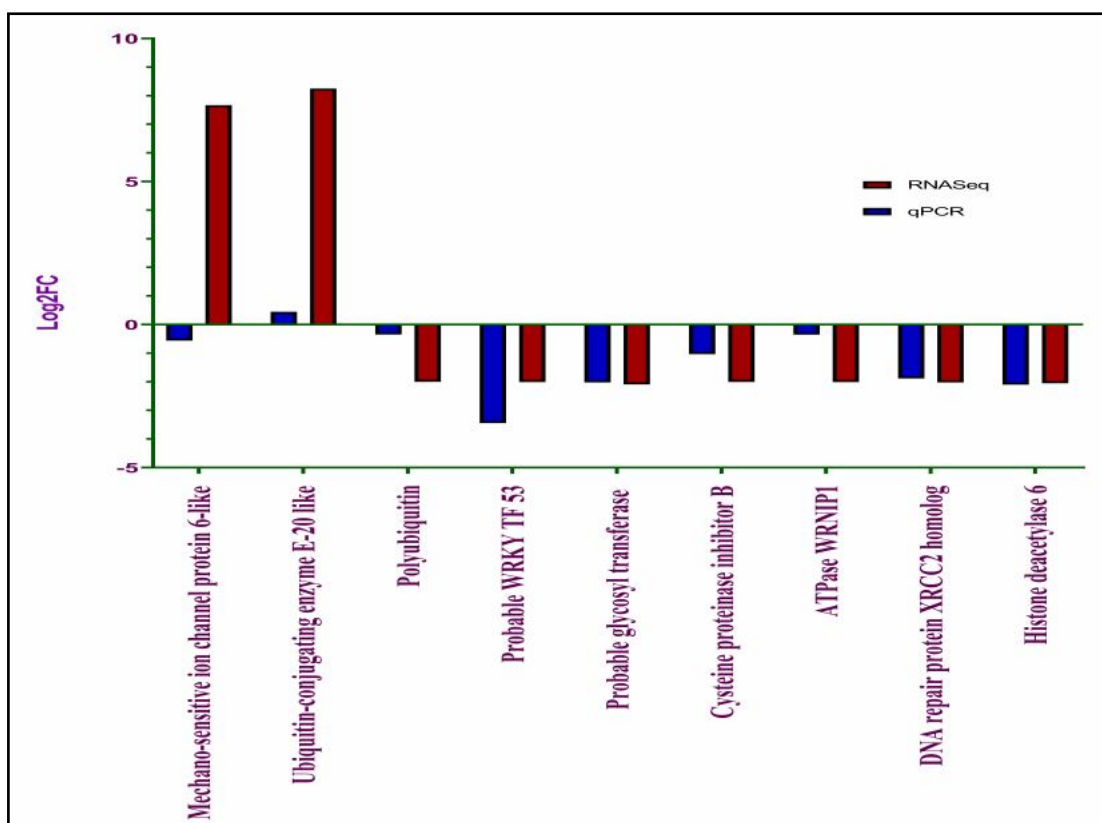


Fig. 29a: RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in storage root transcriptome of temperate and tropical cultivars

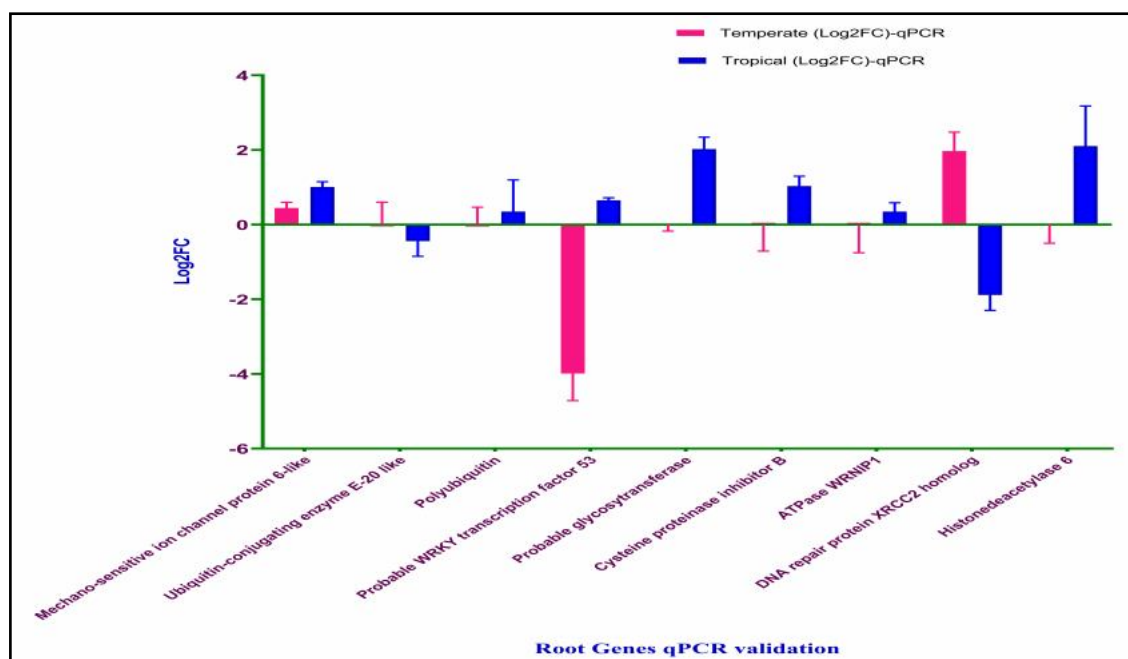


Fig. 29b: RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in storage root transcriptome of temperate and tropical cultivars

4.16.1 Validation of DEGs identified in FPL transcriptome by RNA-Seq results with quantitative real-time PCR (qRT-PCR)

To assess the validity and reliability of our RNA-seq data in identifying flower induction-related DEGs, qRT-PCR was performed using gene-specific primers for all five, of the differentially expressed putative gene specific DEGs for GI (GIGANTIA), CO (CONSTANS); PhyA (PHYTOCHROME A); AP2 (APETALLA2); FKF1: (FLAVIN BINDING KELCH REPEAT, F-BOX-1) between two tropical and temperate carrot FPL tissue transcriptome which included genes encoding proteins with homologs that were previously functionally or non-functionally annotated in *Arabidopsis* and novel genes in carrot (*Daucus carota L.*) like CO and GI along with other (Fig. 30). Among these DEGs, the qRT-PCR profiles of five genes specific DEGs chosen agreed with expression pattern for those obtained from the original RNA-seq results, although the expression fold-change displayed minimal differences (Fig. 30). The expression patterns of these selected DEGs in the tropical and temperate genotypes between RNA-seq and qRT PCR analyses displayed high similarity, thereby confirming the reliability of our RNA-seq data.

4.17 Expression pattern of flowering related DEGs between temperate and tropical cultivars by qPCR

Under phytotron experimental condition the expression pattern of 14 different flowering specific genes was studied across temperate and tropical cultivars. 14 flowering related genes were selected *Arabidopsis* crop as a reference. The following genes were used for qPCR experiment viz., dcSOC1, AP2, FKF, dcCOL2, LEAFY, PHYA, PHYB, GI, FT regulated by photoperiodism and two circadian genes CCA, TOC1 at three time point (8am, 4pm, 8pm) and flowering repressors FLC, SVP, MIR172 were included (Figure 31a). Function of individual gene is clearly mentioned in material methods.

The expression pattern of FT, GI, doSOC1 was upregulated in both temperate and tropical carrots and it was comparatively high in tropical carrots. Whereas, other genes like dcCOL2 were downregulated in their expression and FKF which stabilize CONSTANS (dcCOL2) was also downregulated in expression between temperate and tropical cultivars. Expression of LEAFY was upregulated in tropical cultivar and in

contrast, it was downregulated in temperate cultivar. Similar pattern of gene expression was observed for PHYB and GI. The expression pattern was reverse for AP2. In other words, AP2 expression was upregulated in temperate and downregulated in tropical. AP2-like genes inhibit the onset of flowering by repressing expression of FT and SOC1. All circadian related genes CCA, TOC1 collected at 8am, 4pm and 8pm was also downregulated in expression. PHYA expression was downregulated in both the cultivars, which is known to stabilize the expression of *dcCOL2*. PHYB acts to delay floral transition by destabilizing the CO protein through ubiquitination and expression was downregulated between temperate and tropical cultivars. MIR172 expression was significantly low in temperate. LEAFY transcription factor expression was high in tropical carrot cultivars than temperate carrots. Expression pattern of all the flowering related genes evaluated under phytotron is shown in heat map Fig. 31b.

4.18 Gene Regulatory Network (GRN) analysis for flower transition between temperate and tropical carrots:

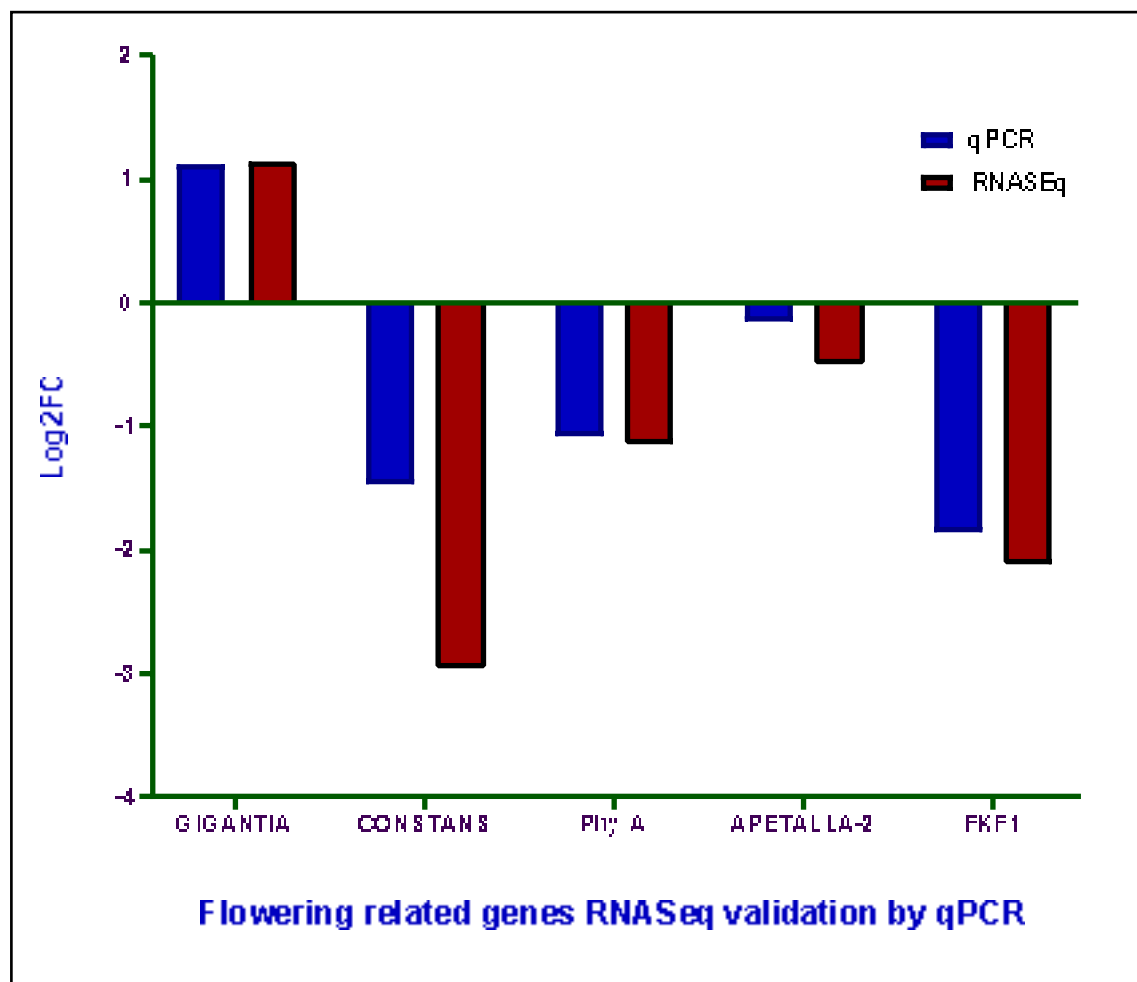
GRN clearly shows the interaction between COL, COL1, COL2 with FKF, DOF, GI and GAI mediated by circadian gene TOC1 and SPL4 and SPL5, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* floral activators, which induce the expression of FT by targeting repressor *miRNA156*. Expression of SPL upregulate the level of *miRNA172* and induce flowering in carrot (Fig. 33).

4.19 Assessment of stability, adaptability and yield performance of carrot (*Daucus carota sub spp. sativus*) cultivars in North Karnataka (tropical region)

Stability analysis for below mentioned seven superior tropical and temperate adapted genotypes across three locations (Bagalkote, Arabhavi, Devihosur) and three seasons at Bagalkote (year 2017, 2018 and 2019) was carried out at vegetative phase to compare secondary storage root performance across seasons and environment. The results are presented below.

Stability analysis across seasons:

Analysis of variance revealed that mean sum of squares due to genotypes across the seasons (Table 33a) were significant for all the seven traits viz., shoot length, shoot



GI: GIGANTIA; CO-CONSTANS; PhyA-PHYTOCHROME A; AP2-APETALLA2; FKF1: FLAVIN BINDING KELCH REPEAT F-BOX-1

Fig. 30: RNA-Seq differentially expressed genes data validation by quantitative real-time PCR (qRT-PCR) in FPL tissue transcriptome of temperate and tropical cultivars

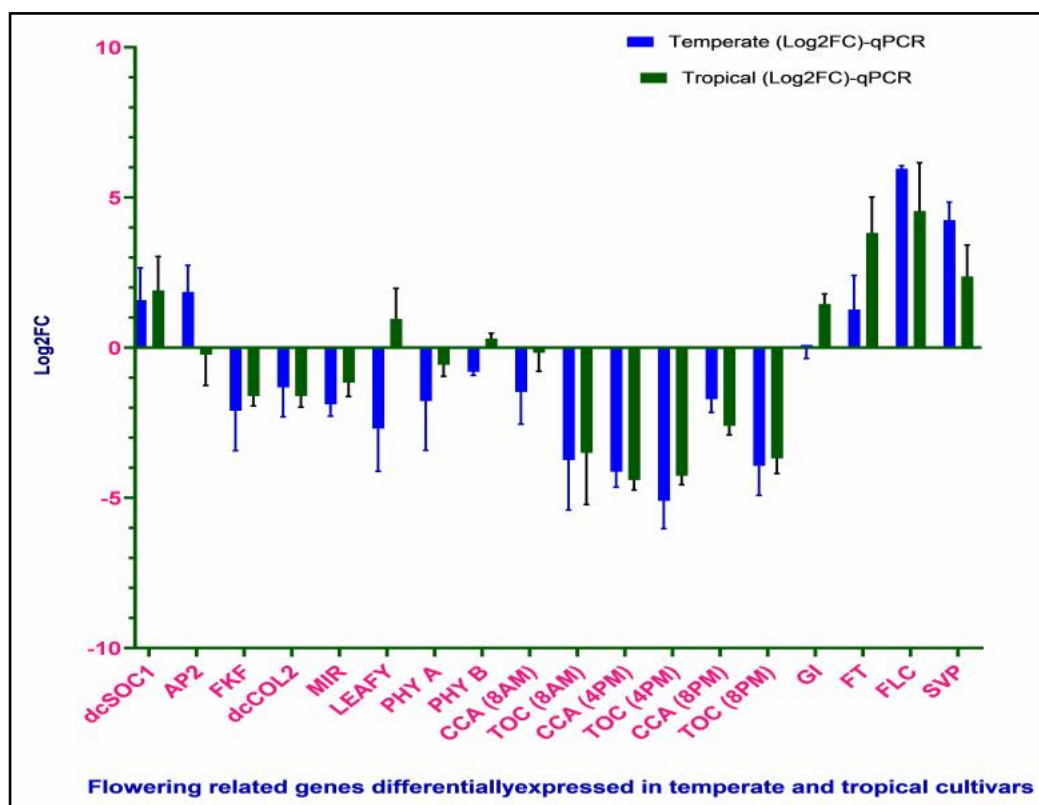


Fig. 31: qRT PCR validation of flowering related DEGs between temperate and tropical cultivars based on gene expression values

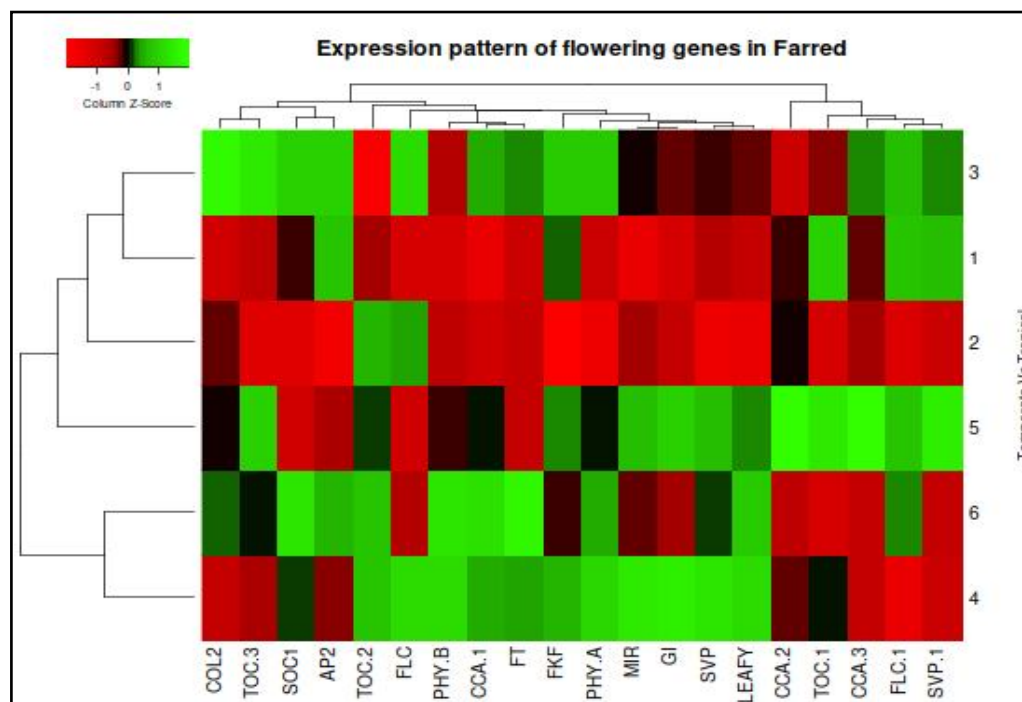


Fig. 32: Heat map showing expression pattern of flowering related genes between temperate and tropical cultivars. CCA1, TOC1 -8am samples, CCA2, TOC2 4pm samples, CCA3, TOC3- 8pm samples

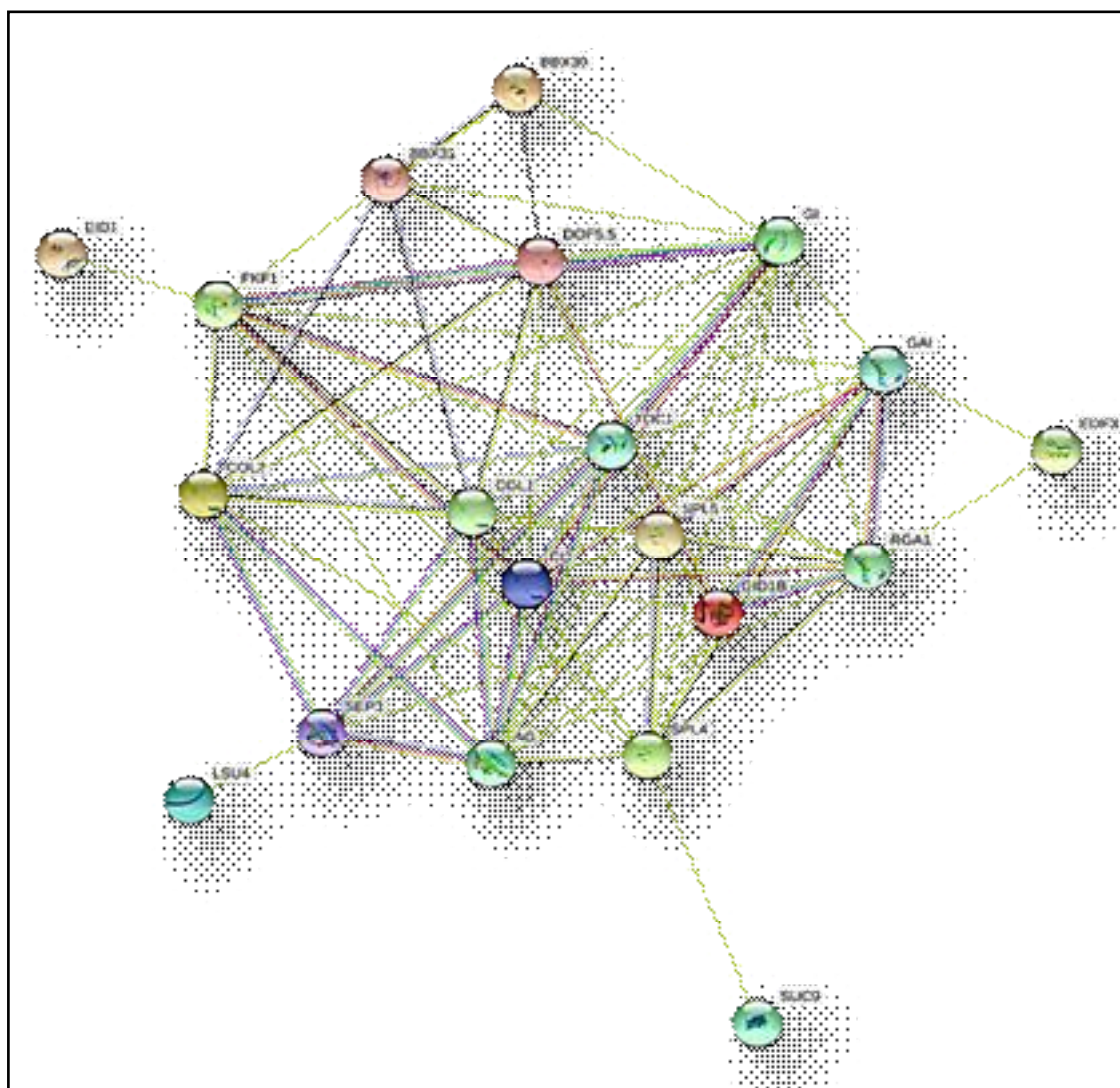


Fig. 33: Identification of conserved network of flowering regulation genes critical for floral transition in carrot identified from *A. thaliana* orthology and network drawn in STRING v. 11

weight, shoulder width, number of petioles, root length, root width and root weight indicating the presence of genetic variability among the genotypes involved in the study for these traits.

The pooled analysis of variance for various traits in carrot over three seasonal conditions (Table 33a) revealed that the variation due to $G \times E$ interaction has been partitioned into two, the predictable component due to linear regression and the unpredictable one due to pooled deviations from regression. Mean sum of squares due to Environment + (Genotype $\times$ Environment) was observed significant for shoot weight (g), shoulder width (mm), root length (cm) which depicted the existence of genotype $\times$ environment interaction. The linear contribution of environment on the performance of genotypes was significant for almost all the traits under study except for shoot length and number of petioles indicating that environmental effects were predominant.

The mean square due to genotype $\times$ environment (Linear) when tested against pooled deviation, were significant for shoot weight (g). This indicated significant rate of linear response of the genotypes to environmental changes for this trait. Non-significant effect of genotype $\times$ environment (linear) for shoot weight (g), shoulder width (g), number of petioles, root length (cm), root weight (g) and root width (mm) indicated that the different genotypes did not differ genetically in their response to different environments in seasonal conditions. The pooled deviation when tested against pooled error was found significant for shoot length (cm), root length (cm) and root width (mm) under studied indicating the important contribution of non-predictable component.

The mean square due to environments (Lin) were significant for all the traits namely shoot length, shoot weight (g), shoulder width (g), root length (cm), root width (mm), root weight (g) which indicated the presence of variable environments in expression of all the traits across seasons (Table 33a). Significant environment mean sum of squares were observed across the seasons for all the three productivity traits and number of petioles. The presence of genotypes $\times$ environment interaction was also significant for shoot weight whereas, it was found non-significant for shoot length (cm), shoulder width (mm), number of petioles, root length (cm), root width (mm) and root weight (g) that indicated the differential response of genotype to all seasonal conditions.

Table 33a: Mean Squares due to different source of variation for various quality traits in carrot (*Daucus carota* L)- ANOVA for Eberhart and Russel Model-Across seasons

Source of Variation	DF	Shoot length	Shoot weight	Shoulder width	Number of petioles	Root length	Root width	Root weight
Replication within Environment	6	19.552	493.104*	4.065	0.617	1.772	19.268	133.806
Genotypes	6	302.499**	2822.874** *	85.523***	1.016	14.47	25.515	605.795*
Environment+ (Genotype × Environment)	14	38.522	824.704**	103.668***	0.871	32.994*	45.804	378.82
Environments	2	3.27	1897.922** *	702.296***	1.06	173.458* **	122.207	1246.041*
Genotype × Environment	12	44.397	645.834**	3.897	0.84	9.583	33.07	234.283
Environments (Lin.)	1	6.539	3795.844** *	1404.592***	2.12	346.917* **	244.415*	2492.082**
Genotype × Environment (Linear)	6	50.387	1206.584** *	5.133	1.196	12.738	21.905	296.989
Pooled Deviation	7	32.920*	72.93	2.28	0.415	5.510***	37.916**	147.067
Pooled Error	36	13.125	184.71	3.543	0.432	1.128	9.091	99.868
Total	20	117.715	1424.155	98.224	0.915	27.437	39.718	446.913

* Significant at 5% level of significance,** Significant at 1% level of significance

The mean performance, regression (b_i) and squared deviation (s^2d_i) for seven morphological traits are presented in Table 33b across seasons. According to Eberhart & Russell (1966) model a stable genotype has high mean yield, $b_i = 1$ and $S^2d_i = 0$.

For shoot length, genotypes UHSBC_117 (G3) and UHSBC_67 (G4) and UHSBC_34-1 (G6) had higher mean than overall mean, b_i significantly greater than 1, non-significant S^2d_i . Therefore, UHSBC_117 (G3) and UHSBC_67 (G4) and UHSBC_34-1 (G6) are stable and well adaptable to favourable environment. Genotypes UHSBC_32-2 (G1), UHSBC_17 (G2), UHSBC_23-1 (G5), UHSBC_100 (G7) was found unstable due to their significant S^2d_i values.

For number of petioles, UHSBC_32-2 (G1), UHSBC_17 (G2) and UHSBC_34-1 (G6) had regression coefficient greater than unity ($b_i > 1$) with non-significant deviation from regression line which indicates its adaptation to favourable environment. UHSBC_117 (G3), UHSBC_67 (G4), UHSBC_23-1 (G5) and UHSBC_100 (G7) showed regression coefficient lesser than unity ($b_i < 1$) with non-significant deviation from regression line so are suitable under unfavourable environment. Genotype UHSBC_32-2 (G1) and UHSBC_34-1 (G6) which are comparable to average mean showed regression coefficient greater than one ($b_i > 1$) with non-significant deviation from regression line reflecting its preference under favourable environments.

For shoot weight, genotypes UHSBC_67 (G4), UHSBC_23-1 (G5), UHSBC_34-1 (G6) displayed unit regression coefficient ($b_i = 1$) and non-significant S^2d_i . Therefore, they are regarded as stable genotypes across all seasons. UHSBC_100 (G7) showed regression coefficient less than unity ($b_i < 1$) with non-significant S^2d_i . UHSBC_17 showed ($b_i > 1$) with non-significant deviation from regression line showing its preference under favourable environments. Hence, they are suitable under unfavourable environments. UHSBC_32-2 (G1), UHSBC_117 (G3) had significant S^2d_i values and found unstable.

For shoulder width, genotypes UHSBC_32-2 (G1), UHSBC_17 (G2), UHSBC_23-1 (G5) and UHSBC_100 (G7) had $b_i > 1$ and non-significant S^2d_i indicating adaptable to favourable environment. UHSBC_117 (G3), UHSBC_67 (G4) and UHSBC_34-1 (G6) had significant positive S^2d_i and found unstable across seasons in varying environment.

For root length, UHSBC_117 genotype was superior over average mean, showed regression coefficient $b_i > 1$ with non-significant deviation from regression line (S^2_{di}) reflecting its stability over favourable environment. Whereas, remaining all genotypes showed significant S^2_{di} and found unstable across seasons.

For root width, UHSBC_117 (G3) and UHSBC_34-1 (G6) showed superior yield and with regression coefficient more than one ($b_i > 1$) with non-significant deviation from regression line indicates its adaptation to favourable environments. UHSBC_100 had $b_i < 1$ with non-significant S^2_{di} , so are suitable under unfavourable environments. Remaining genotypes were considered to be unstable as they had significant S^2_{di} .

For root weight, UHSBC_32-2 (G1), UHSBC_23-1 (G5) and UHSBC_100 (G7) exhibited superior performance coupled with regression coefficients lesser than unity ($b_i < 1$) with non-significant deviation from regression line so are suitable under unfavourable environment. However, UHSBC_117 (G3), UHSBC_67 (G4) had ($b_i > 1$) with non-significant deviation from regression line reflecting its adaptation under favourable environments. However, UHSBC_17 (G2) and UHSBC_34-1 (G6) possess regression co-efficient close to unity but display significant S^2_{di} . Hence, are regarded as unstable genotypes across seasons.

Character wise grouping of genotype for stability analysis under different environments across seasons (Table 33c), clearly showed that none of the genotypes were found adaptable across seasons for all the seven traits evaluated under the present study. Only for shoot weight UHSBC_67 (G4), UHSBC_23-1 (G5), UHSBC_34-1 (G6) genotypes, displayed unit regression coefficient ($b_i = 1$) and non-significant S^2_{di} . Thus, these genotypes can be used for general cultivation across all environments. UHSBC_100 (G7), exhibited non-significant deviation from regression and regression coefficient less than unity ($b_i < 1$) with mean value close to the population for shoot weight, number of petioles, root width and root weight, indicating its stability under unfavourable environments. Genotype UHSBC_23-1(G5) was found stable for number of petioles, root weight across unfavourable environments. UHSBC_117 (G3) for shoot length, root length, root width and, root weight had stability under favourable environments showing non-significant deviation from regression and regression coefficient around unity ($b_i > 1$) along with mean value close to the population mean.

Table 33b: Mean value, regression coefficient (β_i) and variation due to deviation ($\sigma^2 di$) for seven carrot genotypes across seasons

Sl. No.	Genotypes	Shoot length			Number of petioles			Shoot weight			Shoulder width		
		μ Mean	β_i	$\sigma^2 di$	μ Mean	β_i	$\sigma^2 di$	μ Mean	β_i	$\sigma^2 di$	μ Mean	β_i	$\sigma^2 di$
1	UHSBC_32-2	68.609	0.62	25.67	82.776	2.76*	-78.25	28.533	0.91	0.66	7.808	2.39	-0.46
2	UHSBC_17	68.092	-2.59	56.25*	118.589	1.94*	-99.62	32.061	1.27	-1.86	8.583	1.74	-0.46
3	UHSBC_117	66.35	1.37	-13.87	97.6	0.43	-221.92	28.654	0.9	5.08	8.511	-0.52	0.03
4	UHSBC_67	63.962	1.31	-7.19	70.614	-0.86	-167.95	24.477	1.06	-3.62	7.949	1.062*	0.53
5	UHSBC_23-1	59.454	13.08	23.3	74.459	-0.14	-212.07	25.733	1.06	-3.19	7.56	1.25	-0.45
6	UHSBC_34-1	60.911	4.53	-10.53	104.148	2.83	-99.41	28.223	1.03	-2.89	9.089	1.13	0.72
7	UHSBC_100	39.622	-11.33	58.52*	23.806	0.04	-211.64	15.339	0.76	-3.54	7.578	3.53	-0.23
	Population Mean	61			81.713			26.146			8.154		

Sl. No	Genotypes	Root length			Root width			Root weight		
		μ Mean	β_i	$\sigma^2 di$	μ Mean	β_i	$\sigma^2 di$	μ Mean	β_i	$\sigma^2 di$
1	UHSBC_32-2	18.8	1.16	3.5	31.8	2.3	112.86**	89.267	-0.14	-28.31
2	UHSBC_17	18.392	1.43	0.05	27.847	0.75	50.64*	72.608	1.48	124.03
3	UHSBC_117	16.417	1.26	-0.41	30.038	1.27	-3.54	72.744	2.5	-90.16
4	UHSBC_67	16.632	1.33	0.27	26.867	1.1	30.99	60.8	1.26	-62.08
5	UHSBC_23-1	19.878	0.91	21.02***	26.241	0.12	21.71	74.34	0.62	-77.62
6	UHSBC_34-1	18.133	0.98	1.41	33.247	1.45**	-10.54	63.956	1.26	534.14*
7	UHSBC_100	13.183	-0.07	4.19*	25.712	0.02*	-10.52	43.356	0.02*	-103.55
	Population Mean	17.348			28.822			68.153		

*Significant at 5% level of significance, ** Significant at 1% level of significance

Table 33c: Character wise classification of genotype for stability and adaptability under different environments across seasons

Characters	Poor/Unfavourable environments bi <1	Rich/favourable environments bi>1	Overall environments bi=1
Shoot length	-	UHSBC_117 (G3) and UHSBC_67 (G4), UHSBC_34-1 (G6)	
Number of petioles	UHSBC_117 (G3), UHSBC_67 (G4), UHSBC_23-1 (G5), UHSBC_100 (G7)	UHSBC_32-2 (G1), UHSBC_17 (G2), UHSBC_34-1 (G6)	-
Shoot weight	UHSBC_100 (G7)	UHSBC_17 (G2)	UHSBC_67 (G4), UHSBC_23-1 (G5), UHSBC_34-1 (G6)
Shoulder width	-	UHSBC_32-2 (G1), UHSBC_100 (G7) UHSBC_17 (G2), UHSBC_23-1 (G5)	-
Root length	-	UHSBC_117	-
Root width	UHSBC_100	UHSBC_117 (G3), UHSBC_34-1 (G6)	-
Root weight	UHSBC_32-2 (G1), UHSBC_23-1 (G5), UHSBC_100 (G7)	UHSBC_117 (G3), UHSBC_67 (G4)	-

Stability analysis across environments:

The combined analysis of variance across three environments means sum of squares due to genotypes (Table 34a) was significant for shoot length, Number of petioles, shoot weight, and shoulder width, root length and root width traits indicating the presence of genetic variability among the genotypes involved in the study. Therefore, partition of these characters was done as per Eberhart and Russell model (1966) in order to know magnitude of linear and non-linear components of variation. The environment and GEI component can also be subdivided into environment (linear), $G \times E$ (linear) and pooled deviations from regression. Variation due to $G \times E$ interaction has been partitioned in such a way that, predictable component is due to linear regression and the unpredictable one is due to pooled deviations from regression.

The three stability parameters viz., mean (μ), regression coefficient (β_i) and deviation from linearity (σ^2_{di}) were estimated for seven characters and the result obtained are presented character wise in table 34b. According to Eberhart and Russel (1966) the desirable and stable genotypes is one with the high mean yield, β_i close to unity and low and non-significant σ^2_{di} . However, the condition may vary with respect to selection of characters.

For shoot length, genotypes UHSBC-34-1 (G6) and UHSBC_67 (G4), had higher mean yield, unit regression coefficient ($\beta_i=1$) and non-significant σ^2_{di} . Thus, they were stable, high yielding genotypes which can be adapted to all the environments. Genotypes UHSBC_23-1 (G5) and UHSBC_117 (G3) had higher mean than overall mean, β_i significantly greater than 1, non-significant σ^2_{di} . Therefore, UHSBC_23-1 (G5) and UHSBC_117 (G3) are stable and well adaptable to favourable environments. UHSBC_100 (G7) exhibited non-significant deviation from regression and regression coefficient value less than unity ($\beta_i < 1$) with mean value lower than the population mean indicating their stability under unfavourable environments. Genotypes, UHSBC_32-2 (G1), UHSBC_17 (G2) were found unstable due to their significant σ^2_{di} values.

Table 34a: Mean Squares due to different source of variation for various quality traits in carrot (*Daucus carota* L)- ANOVA for Eberhart and Russel Model across locations

Source of Variation	DF	Shoot length	Shoot weight	Shoulder width	Number of petioles	Root length	Root width	Root weight
Replication within Environment	6	15.103	0.785	419.146**	10.587	3.349*	7.471	262.667
Genotypes	6	526.415***	10.305***	2913.404***	110.161***	1.523	10.502	136.831
Environment+ (Genotype × Environment)	14	60.598*	7.766***	1418.167***	8.354	39.012***	8.129	168.859
Environments	2	324.700***	36.058***	6763.322***	6.68	268.798***	26.355*	133.756
Genotype × Environment	12	16.582	3.051*	527.308**	8.633	0.714	5.091	174.71
Environments (Lin.)	1	649.401***	72.115***	13526.640***	13.36	537.597***	52.710**	267.513
Genotype × Environment (Linear)	6	18.936	5.471**	1000.664***	12.843	0.726	6.652	121.578
Pooled Deviation	7	12.194	0.541	46.244	3.791	0.601	3.026	195.292
Pooled Error	36	12.571	1.458	274.505	5.28	2.297	5.557	127.261
Total	20	200.343	8.528	1866.738	38.896	27.765	8.841	159.251

*Significant at 5% level of significance,** significant at 1% level of significance

For number of petioles, UHSBC_117 (G3) and UHSBC_67 (G4) had higher mean yield and regression co-efficient around unity and non-significant S^2_{di} . Thus, are found to be adaptable to all the environments across environments. UHSBC_17 (G2) had regression coefficient greater than unity with non-significant deviation from regression line and higher overall population mean which indicates its adaptation to favourable environment. UHSBC_23-1 (G5), UHSBC_34-1 (G6), UHSBC_100 (G7) genotypes showed regression coefficient lesser than unity ($b_i < 1$) with non-significant deviation from regression line so are suitable under unfavourable environment. Genotype UHSBC_32-2 (G1) and which are comparable to higher population mean showed regression coefficient greater than one ($b_i > 1$) with significant deviation from regression line, hence found unstable due to their significant S^2_{di} values.

For shoot weight, genotypes UHSBC_17 (G2) presented unit regression coefficient ($b_i = 1$) and non-significant S^2_{di} . Therefore, it is regarded as stable genotype across all environments. UHSBC_32-2 (G1) and UHSBC_67 (G4) had regression coefficient greater than unity with non-significant deviation from regression line and higher overall population mean which indicates its adaptation to favourable environment. UHSBC_117 (G3), UHSBC_23-1 (G5), UHSBC_34-1 and (G6) UHSBC_100 (G7) showed regression coefficient less than unity ($b_i < 1$) with non-significant S^2_{di} . Hence, are suitable under unfavourable environments. None of the genotypes had significant S^2_{di} values for shoot weight, therefore no genotypes were found unstable.

For shoulder width, genotypes UHSBC_117 (G3), UHSBC_67 (G4), UHSBC_23-1 (G5) and UHSBC_100 (G7) had $b_i > 1$ and non-significant S^2_{di} indicating adaptation to favourable environment. UHSBC_17 (G2) and showed regression coefficient ($b_i = 1$) significant S^2_{di} and regarded as unstable genotype across environments. Genotypes UHSBC_32-2 (G1), UHSBC_34-1 (G6) showed regression coefficient less than unity ($b_i < 1$) with non-significant S^2_{di} . Hence, are suitable under unfavourable environments.

For root length, UHSBC_32-2 (G1), UHSBC_17 (G2), UHSBC_117 (G3), UHSBC_67 (G4), and UHSBC_23-1 (G5) genotypes were superior over average mean, showed regression coefficient close to one with non-significant deviation from regression line (S^2_{di}) reflecting its stability over changing environments. Genotype UHSBC_100 (G7) had $b_i > 1$ and non-significant S^2_{di} indicating its adaptability to favourable environment. Genotypes UHSBC_32-2 (G1), UHSBC_34-1 (G6), showed regression coefficient less than unity ($b_i < 1$) with non-significant S^2_{di} . Hence, UHSBC_34-1 (G6) is suitable under unfavourable environments. None of the genotypes evaluated in the study showed significant S^2_{di} for root length and found unstable across environments.

For root width, UHSBC_100 (G7), showed superior population mean and with regression coefficient close to one with non-significant deviation from regression line indicates adaptation to all the environments. UHSBC_67 (G4), had $b_i > 1$ and non-significant S^2_{di} indicating adaptable to favourable environment. UHSBC_17 (G2), UHSBC_117 (G3), UHSBC_34-1 (G6), had $b_i < 1$ with non-significant S^2_{di} , so are suitable under unfavourable environments. Remaining UHSBC_32-2 (G1), UHSBC_23-1 (G5), genotypes were considered to be unstable as they had significant S^2_{di} .

For root weight, none of the genotypes showed unit regression coefficient ($b_i = 1$) and non-significant S^2_{di} and found adaptable to all the environments. UHSBC_23-1 (G5), UHSBC_34-1 (G6) and UHSBC_100 (G7) genotypes exhibited superior performance coupled with regression coefficients lesser than unity ($b_i < 1$) with non-significant deviation from regression line so are suitable under unfavourable environment. Yet, UHSBC_17 (G2), and UHSBC_117 (G3) had ($b_i > 1$) with non-significant deviation from regression line reflecting its adaptation under favourable environments. However, UHSBC_32-2 (G1) and UHSBC_67 (G4) possess regression co-efficient more than one and display significant S^2_{di} . Hence, are regarded as unstable across environments.

Character wise classification of genotypes used in the present investigation (Table 34c) clearly revealed that, UHSBC_34-1 (G6) exhibited non-significant deviation from regression and regression coefficient less than unity ($b_i < 1$) with mean value close to the population for all traits except shoot length indicating its stability under unfavourable environments. Whereas, UHSBC_100(G7), showed stable performance across all environments for shoot length, number of petioles, shoot weight, root weight. Genotype UHSBC_23-1(G5), was found stable for number of petioles, shoot weight and root weight across all environments. Thus, these genotypes can be used for general cultivation across all environments. UHSBC_117(G3), for shoot length, shoulder width and root weight and UHSBC_67(G4), for shoot weight, shoulder width and root weight traits, showing stability under favourable environments showing non-significant deviation from regression and regression coefficient around unity ($b_i > 1$) along with mean value close to the population mean.

For root length, genotypes such as, UHSBC_32-2(G1), UHSBC_17(G2), UHSBC_117(G3), UHSBC_67(G4) and UHSBC_23-1(G5), were having unit regression coefficient ($b_i \approx 1$) and deviation from regression non-significantly deviate from zero ($S^2_{di}=0$) with high mean performance over the population mean across all environments, thereby indicating their stability and adaptability under all environments. Data further revealed that, UHSBC_67 (G4) genotype had stability for shoot length and number of petioles as well across three different environments.

The environmental index (EI) was also estimated for the three root productivity traits, across locations and seasons, results revealed that, Location-1 (Bagalkote) showed positive for root length with four genotypes showing positive phenotypic index (PI) viz., UHSBC_67(G4), UHSBC_17(G2), UHSBC_23-1(G5) and UHSBC_117 (G3)). But, other two locations (Location-2 and Location-3) showed poor performance for root length showing negative EI. For root weight also Location-1 (Bagalkote) and Devihosur (Location-3) showed positive EI. But, Arabhavi (Location-2) location showed negative EI. Among the genotypes, four genotypes showed positive PI (UHSBC_17(G2), UHSBC_67 (G4), UHSBC_117(G3) and UHSBC_34-1 (G6)). In

general, among the three locations, Bagalkot location was found to be more favourable than other two locations, and genotypes viz., UHSBC-17 (G2), UHSBC-67 (G4), UHSBC-117(G3) and UHSBC_34-1 (G6) found outperforming over other genotypes as presented in Table 35.

While, among seasons, year 2017 showed positive EI for root length and root weight genotypes UHSBC_32-2(G1), UHSBC_17 (G2) and UHSBC_23-1(G5) showed positive phenotypic index indicating their superiority over other genotypes across seasons.

Table 34b: Mean value, regression coefficient (bi) and variation due to deviation (s^2di) for seven carrot genotypes across locations

Sl. No.	Genotypes	Shoot length			Number of petioles			Shoot weight			Shoulder width		
		μ Mean	β_i	S^2di	μ Mean	β_i	S^2di	μ Mean	β_i	S^2di	μ Mean	β_i	S^2di
1	UHSBC_32-2	64.044	0.99	30.4	12.444	2.24	0.24	126.02	2.07	-244.48	37.349	-3.214	-5.99
2	UHSBC_17	61.444	1.15	1.31	10.356	1.64	-0.78	107.507	1.09	-293.43	43.057	0.34	11.2
3	UHSBC_117	61.741	1.54	-12.68	9.157	1.03	-1.14	82.827	0.75	-177.22	39.001	1.22	-3.05
4	UHSBC_67	58.711	0.98	-1.49	8.956	0.91	-0.45	83.343	1.78	-165.1	39.856	5.31	-0.3
5	UHSBC_23-1	56.778	1.2	-11.82	9.578	0.58	-1.34	77.596	0.85	-293.37	39.576	2	-5.93
6	UHSBC_34-1	62.467	1.07	-5.38	10.022	0.49	-0.92	104.542	0.44	-284.55	39.053	-0.41	-5.89
7	UHSBC_100	26.407	0.07	-5.51	6.256	0.12*	-1.36	28.204	0.02*	-284.31	24.274	1.76	-5.76
Population Mean		55.942			9.538			87.148			37.452		

*Significant at 5% level of significance, ** Significant at 1% level of significance

Sl. No.	Genotypes	Root length			Root width			Root weight		
		μ Mean	β_i	S^2di	μ Mean	β_i	S^2di	μ Mean	β_i	S^2di
1	UHSBC_32-2	20.744	0.97	-2.31	25.691	0.71	2.63	90.731	2.05	115.52
2	UHSBC_17	22.178	0.96	-1.51	27.06	0.53	-4.93	94.491	2.84	-50.25
3	UHSBC_117	20.489	0.97	-2.42	24.537	0.03	-5.78	85.804	1.69	-146.01
4	UHSBC_67	21.862	1.01	-2.37	21.795	2.66	-4.05	90.266	2.6	642.01*
5	UHSBC_23-1	20.516	1.02	-1.69	27.365	1.88	0.75	87.931	-0.5	-40.02
6	UHSBC_34-1	20.667	0.87	-2.03	25.858	0.28	-3.78	87.118	-1.97	-137.76
7	UHSBC_100	21.644	1.19	-0.59	26.028	0.91	-4.48	73.191	0.29	-42.68
Population Mean		21.157			25.476			87.076		

*Significant at 5% level of significance, ** Significant at 1% level of significance

Table 34c: Character wise classification of genotype for stability and adaptability under different environment

Characters	Poor/Unfavourable environments bi <1	Rich/favourable environments bi>1	Overall environments bi=1
Shoot length	UHSBC_100 (G7)	UHSBC_23-1 (G5), UHSBC_117 (G3)	UHSBC_34-1 (G6), UHSBC_67 (G4)
Number of petioles	UHSBC_23-1 (G5), UHSBC_34-1 (G6), UHSBC_100 (G7)	UHSBC_17 (G2)	UHSBC_117 (G3), UHSBC_67 (G4)
Shoot weight	UHSBC_117 (G3), UHSBC_23-1 (G5), UHSBC_34-1 (G6), UHSBC_100 (G7)	UHSBC_32-2 (G1), UHSBC_67 (G4)	UHSBC_17 (G2)
Shoulder width	UHSBC_32-2 (G1), UHSBC_34-1 (G6)	UHSBC_117 (G3), UHSBC_67 (G4), UHSBC_23-1 (G5), UHSBC_100 (G7)	-
Root length	UHSBC_32-2 (G1), UHSBC_34-1 (G6)	UHSBC_100 (G7)	UHSBC_32-2 (G1), UHSBC_17 (G2), UHSBC_117 (G3), UHSBC_67 (G4), UHSBC_23-1 (G5)
Root width	UHSBC_17 (G2), UHSBC_117 (G3), UHSBC_34-1 (G6)	UHSBC_67 (G4)	UHSBC_100 (G7)
Root weight	UHSBC_23-1 (G5), UHSBC_34-1 (G6) UHSBC_100 (G7)	UHSBC_17 (G2), UHSBC_117 (G3)	-

Table 35a: Environmental index and phenotypic index for root productivity traits across seasons

Environmental Index (EI) and Phenotypic index (PI) across seasons for root productivity traits											
	Root Length						Root width			Genotypes Nea	PI
	Year-2017	Year-2018	Year-2019	Genotypes Nea	PI		Year-2017	Year-2018	Year-2019		
UHSBC_32-2	25.67	15.67	15.07	18.80	1.45	UHSBC_32-2	19.00	43.77	32.63	31.80	2.98
UHSBC_17	26.60	13.58	14.99	18.39	1.04	UHSBC_17	25.72	22.83	34.99	27.85	-0.97
UHSBC_67	23.67	12.10	13.48	16.42	-0.93	UHSBC_67	24.55	29.80	35.76	30.04	1.22
UHSBC_117	24.33	12.24	13.32	16.63	-0.72	UHSBC_117	22.84	23.43	34.33	26.87	-1.95
UHSBC_23-1	24.20	13.51	21.93	19.88	2.53	UHSBC_23-1	24.64	30.84	23.24	26.24	-2.58
UHSBC_34-1	23.93	15.33	15.13	18.13	0.78	UHSBC_34-1	26.44	35.32	37.99	33.25	4.43
UHSBC_100	12.38	12.05	15.12	13.18	-4.17	UHSBC_100	25.61	25.85	25.68	25.71	-3.11
Mean	22.97	13.50	15.58	17.35		Mean	24.11	30.26	32.09	28.82	
EI	5.62	-3.85	-1.77			Eindex	-4.71	1.44	3.27		

Environmental Index (EI) and Phenotypic index (PI) across seasons for root productivity traits					
	Root weight				
	Year-2017	Year-2018	Year-2019	Genotypes Nea	PI
UHSBC_32-2	82.53	89.87	95.40	89.27	21.12
UHSBC_17	78.73	47.73	91.36	72.61	4.46
UHSBC_67	96.27	34.43	87.53	72.74	4.59
UHSBC_117	77.73	43.00	61.67	60.80	-7.35
UHSBC_23-1	77.60	64.10	81.32	74.34	6.19
UHSBC_34-1	92.27	49.33	50.27	63.96	-4.19
UHSBC_100	44.27	43.17	42.63	43.36	-24.79
Mean	78.49	53.09	72.88	68.15	
Eindex	10.34	-15.06	4.73		

Table 35b: Environmental index and phenotypic index for root productivity traits across locations

Environmental Index and Phenotypic index across locations for root productivity traits											
	Root Length						Root width			Genotypes Nea	PI
	Bagalkote	Arabhavi	Devihosur	Genotypes Nea	PI		Bagalkote	Arabhavi	Devihosur		
UHSBC_32-2	24.33	12.24	13.32	16.63	-0.72	UHSBC_32-2	22.84	23.43	34.33	26.87	15.40
UHSBC_ 17	26.60	13.58	14.99	18.39	1.04	UHSBC_ 17	25.72	22.83	34.99	27.85	16.38
UHSBC_67	24.20	13.51	21.93	19.88	2.53	UHSBC_67	24.64	30.84	23.24	26.24	14.77
UHSBC_ 117	25.67	15.67	15.07	18.80	1.45	UHSBC_ 117	19.00	43.77	32.63	31.80	20.33
UHSBC_23-1	23.93	15.33	15.13	18.13	0.78	UHSBC_23-1	26.44	35.32	37.99	33.25	21.78
UHSBC_34-1	23.67	12.10	13.48	16.42	-0.93	UHSBC_34-1	24.55	29.80	35.76	30.04	18.57
UHSBC_ 100	12.38	12.05	15.12	13.18	-4.17	UHSBC_ 100	25.61	25.85	25.68	25.71	14.24
Mean	22.97	13.50	15.58	17.35		Mean	24.11	30.26	32.09	28.82	
EI	5.62	-3.85	-1.77			Eindex	-4.71	1.44	3.27		

Environmental Index and Phenotypic index across locations for root productivity traits					
	Root weight				
	Bagalkote	Arabhavi	Devihosur	Genotypes Nea	PI
UHSBC_32-2	77.73	43.00	61.67	60.80	-7.35
UHSBC_ 17	78.73	47.73	91.36	72.61	4.46
UHSBC_67	77.60	64.10	81.32	74.34	6.19
UHSBC_ 117	82.53	89.87	95.40	89.27	21.12
UHSBC_23-1	92.27	49.33	50.27	63.96	-4.19
UHSBC_34-1	96.27	34.43	87.53	72.74	4.59
UHSBC_ 100	44.27	43.17	42.63	43.36	-24.79
Mean	78.49	53.09	72.88	68.15	0.00
Eindex	10.34	-15.06	4.73		

5. DISCUSSION

The temperate type storage root is basically adapted to cool climate carrying all the domesticated traits like uniform root colour, shape, size, and flavour. It is typically characterised by orange flesh with self-core, deeply pigmented and juicy roots with increased carotenoid and sugar contents (Rong *et al.*, 2014). These temperate adapted Western carrots were found to be relatively large and unbranched when grown even under tropics in the present study making it a farmer's choice in cultivation practices. Whereas, the Asiatic domesticated orange carrots being adapted to warmer climate, had significantly larger roots, however, less phloem width and more xylem pattern, light orange flesh with orange or yellow core, having meagre carotene content necessitates the attention for crop-improvement. However, its higher sugar content, better keeping quality without rotting, ability to induce flowering with less or no cold-exposure make it a choice for seed production by farmers henceforth, commonly available to consumers of warmer climate.

Considering its vast diversity in cultivar types adaptable to contrasting climates viz., temperate and tropical types from diverse gene pools such as temperate/Western and Asiatic/tropical types, and the plant material was chosen for the study. Our previous study on population structure clearly distinguished the tropical carrot population from temperate gene pools with moderate genetic differentiation (Chaitra *et al.*, 2020). In the present study, major objectives were to understand the transcriptome dynamics of these two economically important cultivars of carrot in terms of overall genetic makeup and richness of each cultivar for nutrition, development and adaptation mechanism of secondary storage root, response to biotic and abiotic stresses under natural conditions without inducing any particular stress. Objective also included understanding the flowering mechanism of these two contrasting cultivar types. To achieve these goals, two economic tissues including edible part (taproot) and vegetative to reproductive transition part (floral primordia & leaf) from two phases of carrot life cycle such as vegetative and reproductive respectively were carefully chosen from temperate and tropical cultivars for the study. This comparative analysis of the two cultivated gene pools (temperate and tropical) from two important tissues of carrot for storage root as well as flowering pathway is the first report in carrot globally.

Previous research work provided insight into storage root development and genetic complexity between wild and domesticated carrots, wherein both the roots have long domestication record and the favourable traits under domestication are extremely different (Machaj *et al.*, 2018). Whereas, our data was derived from two important cultivated types of carrot *viz.* Asiatic domesticated orange (tropical-type) and Western domesticated orange (temperate-type) which have recent domestication history. It is believed that Asiatic type cultivated carrot originated in south Asia (Iorizzo *et al.* 2013, 2016) considered as first domestication event of cultivated carrot. Historical records and experimental results support central Asia as the centre of origin of cultivated carrot (Que *et al.*, 2019). Yet then, comparative research studies on these two cultivar types are scanty both at the phenotypic and molecular level. Hence, the present research was designed to compare the root transcriptome between tropical type carrot and temperate type carrot root and FPL tissues.

5.1 Phenotypic plasticity of temperate and tropical and carrots

The adaptive plasticity of these two cultivars was evaluated by estimation of root, flower and plant-morphological characters in tropical region at Bagalkot having relatively hot and humid temperature with intense sun light at both vegetative and reproductive stages. Tropical carrot showed superior *per se* performance in terms of root weight and length same like temperate carrot. However, tropical carrot can withstand the water scarcity, with no wilting symptoms during vegetative stage. Tropical cultivar showed early flowering behaviour with less or no vernalization requirement than temperate cultivar during reproductive stage. Temperate carrot was superior in terms of attractive dark orange colour, uniform inner core with no clear xylem or phloem patterning and higher beta carotene in contrast to tropical cultivar, but did not switch from vegetative to reproductive stage under tropics.

Understanding the genetic architecture of this plasticity may possibly provide useful breeding targets for crop improvement specifically in tropical or sub-tropical environments that constitute the major part of the cultivation area of globe. Plasticity refers to a plant's ability to adapt to and cope up with changes in its environment is genetically heritable (Pigliucci, 2005), and is recognised as significant source of phenotypic variation and diversity. Elucidating the mechanisms underlying root

plasticity in terms of ‘secondary storage development’ enable us to understand how domesticated temperate-orange storage root developed and evolved compared to least exposed tropical cultivars during domestication. Further it may provide greater opportunity for selection of cultivars carrying the corresponding traits to various biotic and abiotic stresses alongside plasticity in breeding populations (Pigliucci, 2005).

In order to have deeper insight into this adaptive plasticity of tropical and temperate cultivars, well developed matured roots of these cultivars grown under tropics were subjected to RNA-seq by Hiseq Illumina sequencing as discussed below.

5.2 Transcriptome profiling of root and FPL tissues

Storage roots are important edible tissues in both the cultivated types of carrots and are known to store a large amount of carbohydrates, vitamins, minerals and other metabolites in the enlarged taproots that will be utilized for flowering in the second year if not harvested for consumption. Hence, the objective of the study was to have deeper understanding of key genes involved in regulation of storage root development processes, constitutive biotic and abiotic stress responsive genes under natural field conditions and reproductive transition from vegetative phase *etc.* at the transcriptome level. To achieve this, two economically important cultivar types of carrot *viz.*, tropical adapted orange cultivar (UHSBC-23-1) and temperate adapted orange cultivars (UHSBC-100) were subjected to RNA-seq analysis using Hi-seq Illumina-4000 platform. From two economic developmental phases of carrot life cycle, two important tissues were utilized for the study *viz.*, matured taproot at harvest and floral primordia pooled with leaf (hereafter mentioned as FPL for floral primordia +leaf) tissue during the vegetative and reproductive phases respectively.

Availability of high-quality genome sequence of carrot helped us to construct reference-based assembly of six independent libraries, *i.e.*, three from each cultivar with three biological replicates. An average of 92.71 % of the high-quality raw reads (Phred score >25) were retained with an average of 21.77 M cleaned reads per library after filtering and cleaning indicates the sufficient depth of RNA-seq reads for assembly per library that is much higher than Machaj *et al.*, 2018 (average of 18 M reads/library).

In FPL tissue transcriptome also the quantity of data was sufficient to perform all the standard bioinformatics analysis and identification of reliable DEGs. The 86.74 % of the FPL tissue raw data was of high quality with Phred score >30 for analysis with average of 33.73 M trimmed and filtered high quality cleaned reads per library for generating the assembly.

The estimated size of the carrot genome is 480 Mb (Iorizzo *et al.*, 2016) and the recent transcriptome data of carrot studies utilized ca. 92-94% amount of reads per library for assembly (Meng *et al.*, 2020). However, the quantity of reads in our work was more than other previous carrot transcriptome studies (Ma *et al.*, 2018; Machaj *et al.*, 2018). The total reads used in our study (130 million cleaned reads) was much higher than previous studies of carrot.

A) Transcript assembly of root tissues and quality assessment of assembly

For root tissue, assembly was constructed both by ‘reference based’ and ‘*De novo* based’ to explore the existing carrot genome and to identify cultivar specific transcripts respectively in the genetic background of two contrasting cultivars (tropical and temperate) by utilizing available high computational tools. For reference-based assembly, available high-quality genome of carrot was downloaded from NCBI (ASM162521v1) on 14th March 2021. In six libraries of our root transcriptome, TopHat assembly mapping rate ranged from 71.6 to 82.8% with an average of 73.87 % mapping rate. Our reference-guided transcriptome assembly is better than the previous average assembly of 64.73 % ranging from 0.06 to 66.61 % assembly rate per sample in terms of coverage (Meng *et al.*, 2020).

In *de novo*-based assembly for root tissue, all the cleaned reads were combined together to construct assembly by using trinity assembler. The assembly generated 1,89,857 total transcripts and 1,03143 from 207641471 bases. The percent GC was 39.59 %, with an average contig length of 1093.7 bp, median contig length of 623 bp and N50 contig length of 1890 bp. The proportion of unknown nucleotides (N) in cleaned reads was found to be nil (0). The average contig length (960bp) and N50 contig length (1571bp) in our assembly was higher than the *de novo* assembly of Ma *et al.* (2018) indicating the superior quality of assembly.

In case of FPL tissue, *de novo* assembly was performed. However, the size of the raw reads in FPL tissue was higher *i.e.*, 151 bp paired-end reads. Hence, higher configuration was necessary to perform the computational analysis. Therefore, we opted 128 GB RAM with 50 CPU machine having 2500 GB storage to perform the trinity based *de novo* assembly. The number of transcripts (1, 23,912) and unigenes (49,978) obtained in the FPL assembly was comparable with the previous literature (Machaj *et al.*, 2018). However, the number of transcripts was lesser and number of unigenes was higher than the root assembly. The N50 contig was 1766 bp and 1363 bp for trinity assembly and unigenes respectively with 972.8 bp and 841.76 bp average contig length with the average of ~39 % GC content.

The number of transcripts or the average contig length and also the quantum/depth of reads used for construction of *de novo* assembly is much superior than previous reports (Arias *et al.*, 2020; Ma *et al.*, 2018) of transcriptome analysis of carrot. The assembly quality was also confirmed by BUSCO analysis of root (Eudicot's lineage for orthology search across 31 genomes) with 89.08 % completeness (only 2.08 % fragmented and 8.08 % missing) and average of 98.88 % assembly of individual library to the constructed *de novo* assembly. In case of FPL assembly, BUSCO analysis showed 85.38% completeness with only 4.69% fragmented and 9.93 % missing BUSCO groups. Hence, the *de novo* assembly of the present study for both root and FPL tissues is superior compared to previous *de novo*-based transcriptome studies. As expected from the higher duplication event in carrot, the complete duplicated copies were 47.81 % and 34.74 % respectively in root and FPL tissue BUSCO analysis. This is the first report showing the BUSCO groups completeness assessment in carrot transcriptome data.

B) Transcript abundance estimation and Normalization of root tissues

Gene expression is assumed by sum of all its isoforms computed by counting reads per gene analogous to the reference genome annotation used for mapping. Read counts are normalized to minimize the variability introduced by read length bias (Oshlack *et al.*, 2009; Marioni *et al.*, 2008) and fluctuations in the number of reads for

every run (Mortazavi *et al.*, 2008). Because a read may map to multiple genomic loci, the accuracy of gene or transcript expression estimation depends on the ability of a quantification algorithm to resolve the ambiguities from the sequence-mapping step. In addition, a gene may have multiple alternatively spliced transcripts sharing a common set of exons, where a read mapped to the shared exons may belong to any one of the transcripts. Currently, the handling of these ambiguities is resolved by Poisson model-based quantifier Cufflinks, which probabilistically assign multi-mapped reads based on the assumption that reads from genomic loci follow the Poisson distribution. Cufflinks take input read alignments in BAM formats and reference genome annotation files in GTF/GFF or BED format.

Most important step in RNA-seq pipeline is expression normalization. In the statistical dependence between two variables (paired end reads), the levels of one of the variables vary in precisely fixed way as regards to levels of the other variable. The number of reads produced by each transcript is proportional to its abundance but technical variations in library preparations and sequencing and biological variability between replicates of the same experiment can change the abundance estimation. Hence, inter or intra-sample comparisons of RNA-seq expression estimates can only be done after normalization.

Transcriptome reconstruction and abundance estimation of transcripts in each library of three biological replicates from tropical and temperate carrot cultivars was performed in both reference-based and *de novo* assembly. Cufflink's suite v1.3.0 was used in reference-based assembly and RSEM method in *de novo*-based assembly from Trinity 2.8.6 package to facilitate the normalized transcripts for further downstream analysis such as differentially expressed genes (DEGs) identification. Transcripts were quantified in individual libraries and normalization was done for each library to get matrix from six libraries with FPKM and TPM values in reference based and *de novo* assembly respectively that was utilized for DEG analysis. Cufflinks and RSEM are the standard tools for abundance estimation, filtering spurious contigs and normalization of transcripts in RNA-seq analysis (Trapnell *et al.*, 2012; Li and Dewey, 2011).

5.3 Unigenes identification from *de novo* assembly in root and FPL transcriptome

After RSEM method of transcript abundance estimation, matrix was generated for six libraries of two cultivars comprising of three biological replicates in root tissue. After filtering the spurious or redundant and low-quality transcripts from matrix, highly abundant root transcripts were retained in the matrix with normalized transcripts that are present in almost all the libraries along with multiple isoforms per gene. In order to retain the longest isoform per gene also called as ‘unigenes’ CD-HIT-EST analysis was performed for the RSEM matrix file. A total of 44,856 unigenes that represent the ‘longest isoform’ for each abundantly found gene from RSEM analysis in root transcriptome assembly was utilized for DEG identification. After stringent filtering of low-abundant transcripts and longest isoform and CD-HIT-EST analysis, only 44,856 unigenes were retained.

Similarly, in FPL RNA-seq assembly, the redundant transcripts and poor-quality transcripts were filtered with the procedure as mentioned above. The CD-HIT-EST followed by retaining the longest isoform, a total of 49978 unigenes with single longest transcript isoforms per gene was obtained for FPL tissue. These were considered as unigenes and TF identification was performed for unigenes and these unigenes containing contigs were used for DEG identification across cultivars for FPL tissue.

Although the number of unigenes identified in root assembly are lesser compared to 62427 of Ma *et al.*, 2018 root transcriptome, but the data is in accordance with the number of genes identified in the carrot genome project (Iorizzo *et al.*, 2016). Moreover, multiple levels of filtering; retaining only abundant transcripts and CD-HIT-EST analysis are considered to be highly justifiable analysis in our present study for retaining high quality unigenes than retaining higher number of unigenes. The average contig length of 980.07 bp with the N50 contig length of 1843 bp and median contig length of 472 bp was found in unigenes of roots. N50 contig length was 1363bp with an average contig length of 841.76 bp with the higher GC content of 39.70 % in FPL unigenes. The length of unigenes in roots ranged from 200 bp to 26545bp, and the contig length of >2000 bp (2kb) were 6382 in number (14.23%) which is much higher than Ma *et al.*, 2018 where they got 9.74% longer than 2kb Unigenes indicating the

superior assembly quality of our unigenes. The Unigenes file was further used for RSEM matrix construction to identify DEGs.

Total number of assembled transcripts (contigs) is considerably higher than the number of protein-coding genes in well studied plants with similar size genomes such as *Arabidopsis* (35,386), *Medicago truncatula* (62,319), *Ananas comosus* (27,024), and *Populus trichocarpa* (73,013) (Spriggs *et al.*, 2018), suggesting transcript redundancy. Unlike genome guided assemblers, the currently available *de novo* assembly programs are known to generate high level of redundancy. Among the contributors of this redundancy are the sequencing errors and single nucleotide polymorphisms (SNPs) which create mismatches. Accordingly, redundant sequences get generated as the assembly programs fail to consolidate highly similar sequences. This fact is exacerbated with increasing the number of reads used in the transcriptome assembly (Huang *et al.*, 2016).

5.4 Identification of Differentially expressed genes (DEGs)

5.4.1 DEGs in reference-based assembly of root transcriptome

A total of 3544 differentially expressed genes (DEGs) were identified between tropical and temperate cultivar from cuff-diff with the filtration criteria of $P \leq 0.01$ with Log2FC of ≥ 1.0 or ≤ -1 indicating the diversity in genetic makeup of these two economically important cultivar types of carrot with different domestication history. Among these, 1818 genes were upregulated in tropical or in other words downregulated in temperate cultivars, whereas, 1726 genes were downregulated in tropical cultivar, in other words upregulated in temperate cultivar showing the uniqueness of these carrot types for their genetic architecture or genetic differentiation. This is the ever first report of studies on two diverse gene pools of carrot such as temperate and tropical gene pools by NGS tools such as transcriptome profiling. However, previous report based on genome guided *de novo* assembly on temperate orange cultivar (*D. carota subsp sativus* L.) with the wild species (*D. carota subsp commutatus*) by Machaj *et al.*, 2018 identified at three timepoints of root development and identified DEGs between timepoints within a cultivar and between cultivars. Various numbers of DEGs were identified in their study for different timepoints within and between wild and cultivated

carrot cultivars. Among them, 986 DEGs were significant between wild and cultivated carrot at matured stage (Machaj *et al.*, 2018), wherein, 563 were downregulated and 423 were upregulated in wild compared to cultivated carrot. The number of DEGs identified at matured/marketable stage in our study between temperate and tropical cultivars is significantly high in number and hence, the data generated from the present study provide rich genomic resource for deeper biological insight into these two cultivars.

Hence, to have deeper understanding of these two cultivars, DEGs were also identified from *de novo*-based assembly using two standard R-language based bioinformatics tools of Bioconductor viz., DESeq and edgeR as discussed below.

5.4.2 DEGs identified from *de novo*-based assembly for root and FPL tissues

Identification of DEGs was performed from a RSEM matrix generated from 44856 unigenes by DESeq and edgeR (both are R based tools of the Bioconductor package). DESeq identified 2616 total significant DEGs with the criterion of P value ≤ 0.05 , $\text{Log}_2\text{FC} \geq 2$ or ≤ -2 and FDR value of 0.01, whereas edgeR identified 2250 significant DEGs after filtering with the criteria of p value ≤ 0.05 and $\text{Log}_2\text{FC} \geq 2$ or ≤ -2 . Higher stringency was considered for Log_2FC in *de novo* based DEG identification in order to get highly expressed genes. From these two tools, a total of 2239 common and significant DEGs were identified for downstream analysis for deeper biological understanding with annotation, ontology, GRN *etc.* Among common DEGs 1149 were upregulated in temperate (in other words downregulated in tropical cultivar) and 1090 were downregulated in temperate (or upregulated in tropical cultivar) necessitates the understanding and exploration of these cultivars in carrot crop improvement program.

Comparison of DEGs between reference and *de novo* using three different bioinformatics tools (cuff-diff, DESeq and edgeR) involving temperate and tropical cultivars is again the first report in carrot. Identification of DEGs with multiple tools will increase the accuracy and confidence to identify reliable DEGs with minimum false positives. However, DEGs were identified between wild and cultivated carrots for root development or flowering related or use of temperate cultivated carrots for carotenoid accumulation and anthocyanin biosynthesis have been independently studied by different research groups viz., Machaj *et al.*, 2018; Ou *et al.*, 2017; Ma *et al.*, 2018 and Meng *et al.*, 2020 respectively.

Whereas, in FPL tissues, the number of DEGs obtained were much lesser than root tissue. DESeq tool identified quite higher number of DEGs (1746) compared to edgeR (1794). Finally, 1671 common DEGs identified in both the DEG identification tools were filtered and utilized for annotation to ensure the reliability of the results. The depth of the RNAseq in FPL tissue was much higher than the root transcriptome with 33.73 M reads per library. The higher quantity of the data with higher stringency in identification of DEGs indicates the potentiality of the present research work to understand the molecular mechanism governing floral transition. The identified candidate genes need to be further utilized for other economic traits such as biotic, abiotic stress response genes for manipulating the genetic architecture of wider adaptable tropical carrots or cool climate adapted temperate types.

5.5 Functional Annotation

5.5.1 Functional Annotation and biological insights of DEGs in root transcriptome (Reference based)

For DEGs identified in cuff-diff by reference-based assembly, David tool annotated 43.49 % (1541) of 3544 DEGs including 799 tropical upregulated/temperate cultivar downregulated and 742 tropical downregulated/temperate upregulated DEGs. David identified 20 gene clusters with functional gene sets (FGS) in tropical upregulated with highest Enrichment Score (ES) of 3.04 and 13 gene clusters in tropical downregulated annotated gene sets with the highest of 5.87 ES as discussed in results. Among the top enriched FGS of tropical upregulated lists, 13 genes with WRKY transcription factors clustered in genegroup1 showed highest ES, and highest number of genes (222) was clustered in 2nd group with ES 2.43 comprising of multiple functions genes in which majority are involved in innate and adaptive immune response. WRKY transcriptional factors are known for their function in biotic and abiotic stress response (Eulgem *et al.*, 2000) and also involved in root growth and development (Mohanta *et al.*, 2016). Being widely adapted and less exposed to domestication bottleneck, tropical cultivar is expected to withstand biotic and abiotic stresses with better root development (Chaitra *et al.*, 2020). This is supported by six putative disease resistance genes upregulated in tropical cultivar in FGS 14 with ES of 0.699.

Whereas, in tropical downregulated/temperate upregulated lists, Gene group 1 with highest ES (5.87) was recorded wherein the majority of the genes were oxygen evolving enhancer protein 1 (6/8 genes), followed by group2 (ES-4.21) with chlorophyll a-b binding protein group (9/10 genes) and the highest number of genes were in group 2 (152) with multiple functions. In upregulated, gene group 14 comprise of disease resistance genes. These light harvesting chlorophylls a-b binding (LHCB) proteins are the apoproteins of the light harvesting complexes of photosystem-II helps in sunlight absorption and associated with chlorophylls and xanthophylls (Liu *et al.*, 2013) and found upregulated in previous carrot studies involving wild and cultivated temperate orange also (Machaj *et al.*, 2018). Oxygen-evolving enhancer (OEE) protein 1 is extensively involved in the process of evolution of oxygen, photosynthesis and defence response to bacterium and drought stress (Yang *et al.*, 2015; Zadraznik *et al.*, 2019).

5.5.2 Functional annotation of DEGs from root and FPL tissues (*De novo* based)

In root tissue across cultivars, 2239 DEGs identified as common in DESeq and edgeR tools, were subjected to Blast2GO annotation. Blast2GO is a user-friendly intuitive and interactive desktop application and a bioinformatics platform for high-quality functional annotation, visualization and analysis of NGS output datasets (Conesa *et al.*, 2005). 2239 transcript's sequences were first aligned against NCBI non-redundant (NR) protein database using BLASTx and hits rate, GO-mapping, GO-annotation and InterPro scan results were extracted from the Blast2GO. Among 2239 genes subjected to Blast2GO, 641 (28.63%) showed no hits, and 1598 (71.10 %) showed blast hits to at least one sequence of NCBI, all the 2239 (100%) successfully performed InterPro scan. InterPro provides functional analysis of proteins by classifying them into families and predicting functional domains and important sites (Jones, 2014). Annotation was obtained for 1598 with 71.37 % rate of annotation and hence, considered to be highest among the annotation tools as it considers all the species under eudicots lineage available in NCBI database. Among FPL DEG annotation, out of 1671 DEGs, 1184 blasted with at least 10^{-6} e-value 70.85 % rate, all DEGs performed InterPro scan. Blast2GO is considered to be one of the highly efficient annotation tools among number of annotation tools available. Among the annotated sequences, majority of sequences were homologous to genes from *Daucus carota sub spp. sativus* sequences (89.07%) followed by the *Daucus carota* (0.70%) and *Vitis vinifera* (0.70%), *Camellia*

sinensis (0.62%). From the whole genome carrot sequencing project, highest homology was found with *Vitis vinifera* (Grape) and *Camelia sinensis* (Iorizzo *et al.*, 2016; Wang *et al.*, 2020).

In FPL tissue, a total 1184 out of 1671 (70.85%) genes were successfully annotated with blast and 100% Interpro scan. The proportion of annotated genes are comparable with root transcriptome. Hence, Blast2GO is more efficient in annotation of DEGs than other databases. Gene annotation identified 28 reproductive biology related genes among which tropical carrot cultivar had highest genes (23) compared to temperate and the details are discussed in later sections of this chapter. The proportion of annotation is considered to be higher for the organism like carrot where the genome and transcriptome studies are comparatively limited as compared to model organisms. The most important flowering genes were selected for validation across cultivars. GRN was performed with the key flowering genes as discussed below.

5.6 GO classification and gene set enrichment analysis (GSEA)-Root tissues

A GO annotation is a statement about the function of a particular gene and each hit can be assigned with different GO terms. GO provides a hierarchically organised set of thousands of standardized terms of biological processes (BP), molecular functions (MF) and cellular components (CC). Functionally annotated genes were also assigned to KEGG pathway analysis in the present study. KEGG database is a rich resource for understanding high-level functions and utilities of the biological system. The knowledge base-driven pathway analysis not reduces the complexity of functional analysis by grouping thousands of genes into few hundred pathways, but also increases the explanatory power for the experiment by identifying active pathways in different conditions (Du *et al.*, 2016). Detailed GO classification, phylogenetic tree construction for the identified GO terms in BP, CC, MF and for KEGG pathways followed by GSEA analysis (pgea method) were performed for up and downregulated genes using an integrated web browser (iDEP).

5.6.1 GO classification of DEGs for root and FPL tissues (*de novo* based)

GO categorisation in Blast2GO identified 62 BP sub-categories, 16 CC sub-categories and 26 MF sub-categories from 1598 annotations of DEGs in root

transcriptome. The detailed are discussed below. Further to identify over represented and under represented functional categories, enrichment analysis was performed by Fisher's exact test for up and down regulated genes in the background of total DEGs as reference set (Glass and Girvan, 2014). Fisher's Exact Test (FET) is the most common statistic and is used by many researchers wherein, functional categories are enriched in a set of genes based on gene counts. FET assumes that the genes annotated to a functional term are equally likely to be identified. However, as some genes are annotated to many functional terms while others are only annotated to a few, it assumes that genes have no equal likelihood of being identified in the context of gene functional annotations. FET explains how false assumption might alter predictions made in the context of functional enrichment analysis. Functional enrichment analysis often involves comparing a gene set to all the terms in GO, multiple-hypothesis corrections are generally applied to the results of these statistical tests (Khatri and Drăghici, 2005). These corrections decrease the value at which a comparison between a gene set and a GO term should be considered significant. Commonly used multiple-hypothesis corrections include the Bonferroni, Benjamini and the False Discovery Rate (Glass and Girvan, 2014). In the present study FDR of 0.01 was considered for FET analysis.

Detailed GO terms enriched for root and FPL tissue is presented below.

5.6.2 DEGs in Biological process (BP)-Reference based

Among BP terms, GO terms upregulated in tropical cultivars, 'localization' and 'transport' were most represented categories with >90 genes and grouped in a single cluster of phylogenetic trees indicating their relationship in biological function. 'Transport' is associated with directed movement of substances such as macromolecules, small molecules or ions or cellular components into or out of the cells or between cells by some agents such as transporter, or pore proteins. Whereas, 'localization' links to any process in which a cell, a substance or cellular entity like protein is transported, tethered to or else maintained in a specific location. Hence, upregulation of these terms in tropical cultivar indicates the proper biological function of the cells in the root tissues by transporting and locating the substances or macromolecules necessary for its survival or growth and development. 'Nitrogen compound transport', polysaccharide, hemicellulose and xyloglucan metabolic processes' were among the other GO terms identified for BP.

However, the gene set enrichment analysis for the tropical upregulated BP, highly enriched GO terms were specifically identified carbohydrate metabolic process, macromolecule localization, amide & peptide transport *etc* that are related to ‘Localization’ or ‘transport’ or metabolic processes as discussed in previous section. Amides constitutes important class of plant secondary metabolites and are involved in various defence responses against environmental stress, antifungal, antibacterial, insecticidal or herbicidal activities (Kumar *et al.*, 2018). Similarly other transporters like peptide, protein *etc* also play crucial role in transportation and localisation of various macro/micro molecular entities for cell growth and function.

Among the temperate upregulated/tropical downregulated cultivars, 14 GO terms were obtained for BP and 7 GO terms were enriched for BP in GSEA. Photosynthesis process and associated terms (photosynthesis light reaction, light harvesting, photosystem II assembly) were upregulated in ‘temperate cultivar’ as reported in the previous RNA-seq or whole genome analysis (Machaj *et al.*, 2018; Iorizzo *et al.*, 2016). Other than photosynthesis related terms, regulation of cellular/macromolecular/primary or other metabolic process *etc* were clustered in single group in phylogenetic tree and ‘generation of precursor metabolites and energy’ was enriched in GSEA analysis along with cellular/protein or cellular protein containing component assembly.

Generally, in carrot, genes involved in photosynthesis were differentially expressed and highly enriched during storage root development in cultivated orange carrot and repressed in wild carrot (Machaj *et al.*, 2018). Compared to matured roots, young or developing roots are enriched with carotenoid pathway, however in matured carrot roots carotenoid biosynthetic genes are less abundant and are co-expressed with chloroplast genes (Machaj *et al.*, 2018). Higher expression of photosynthesis related genes such as LHC-II in the cultivated carrot is associated with higher carotenoid accumulation in mature roots (Iorizzo *et al.*, 2016). They also found that, genes that are involved in assembly and function of photosystem-I and II and in plastid development were co-expressed with isoprenoid pathway responsible for carotenoid biosynthesis in carrot (Machaj *et al.*, 2018; Iorizzo *et al.*, 2016; Rong *et al.*, 2014). Similar results are found in other root crops which are edible secondary storage roots such as sweet potato, radish, yam, cassava and potato (Hoang *et al.*, 2020). Light is a vital factor in storage

root biomass production. In root crops, the intensity of light levels of photoperiod regulates storage root development.

As per our knowledge, the present study is the first report on tropical orange carrot by transcriptomics. It can be speculated that organisation and morphogenesis of root tissue, number of factors like cellular organisation, metabolic processes or biotic and abiotic stress response factors play crucial role in plant adaptation to varied climatic conditions. These biological processes were major upregulated factors in tropical carrots and may play crucial role in plant adaptation. The results were also confirmed by *de novo* assembly and annotation with higher stringency in DEG identification ($\text{Log}_2\text{FC} \geq 2, \leq -2$ with FDR 0.01 and value 0.01) as discussed below.

5.6.3 DEGs in Biological process (BP) for root transcriptome-De novo based

A total of 62 BP subcategories were identified in Blast2GO, with the number of gene sequences assigned to each category ranging from total of 723 for the term 'Biological process' to 72 for 'small molecule metabolic process'. The highest number of genes identified for 'localisation' and 'transport' in reference-based terms showed only 127 and 123 genes respectively in Blast2GO. Top functional categories under BP term in Blast2GO were 'cellular processes and associated terms, metabolic process and associated terms, 'biosynthetic process & related terms', 'biological regulation', 'gene expression' and 'response to stimulus' terms. Whereas, in FET, overrepresented BP enriched categories in tropical upregulated/temperate downregulated GO terms comprised of 'protein phosphorylation' with 34 genes of test sets and 12 gene sets from reference set. Machaj *et al.*, 2018 also reported the downregulation of protein phosphorylation in temperate cultivated orange cultivar. 'Protein phosphorylation' is a signalling molecule involved in defence mechanism (MAPK) and cellular signal transduction (Tang *et al.*, 2017). Whereas, 'photosynthesis' and 'photosynthesis light reaction' were enriched and over represented in temperate cultivar upregulated categories in Fisher's exact test. These results were consistent in GO analysis or GSEA where in 'photosynthesis and associated terms' were upregulated in reference-based analysis as discussed above and also in previous studies (Iorizzo *et al.*, 2016; Machaj *et al.*, 2018; Rong *et al.*, 2014). The detailed biological functions in root development, biotic and abiotic stress response, reproduction and biosynthetic processes in different

two important cultivars used in the study are elaborated in the later section of the chapter.

5.6.4 DEGs in Biological process (BP) for FPL transcriptome-*De novo* based

A total of 67 subcategories were found for BP in FPL tissue. As expected for the reproductive phase tissue, the number of genes involved in “reproduction” was enriched with highest number of genes (28) in BP. The other enriched terms included were biological process and metabolic process. Further understanding of these genes with model organism like *Arabidopsis* is necessary to elucidate the flowering pathway in temperate and tropical carrot cultivars belonging to biennial and annual growth habit respectively.

5.6.5 DEGs in cellular components (CC) for root tissue-Reference based

The importance cellular component in GO analysis refers to cellular anatomy in which information pertaining to gene product that perform its function in cellular compartments like mitochondria, chloroplast *etc* or in stable macromolecular complexes such as ribosomes. Among the DEGs identified between tropical and temperate cultivars, integral component of the membrane cell wall related structures or cell periphery were upregulated and enriched in tropical cultivars, whereas, photosystems subunit or chloroplast components which are crucial in photosynthetic processes were upregulated as well as enriched in temperate or they were downregulated in tropical cultivars. This indicates that the cell wall of tropical cultivar has developed its outer membrane structure to facilitate adaptive mechanism to various soil or climatic adaptation. However, over the period of time during evolution of temperate carrot, they have developed thin cell wall with soft and smooth texture due to selective forces by human interference for more attractive temperate carrot than course textures tropical cultivar which was clearly depicted in our phenotypic observations in the present and previous studies (Chaitra, 2017). The enriched photosynthetic cellular structures such as chloroplast, plastid, thylakoid membrane, stroma and other photosynthetic cascades in temperate carrots is the clear indication of higher carotenoid accumulation that co-expressed with higher photosynthetic efficiency in temperate types than tropical cultivars.

It was interesting to note that, despite higher stringency in DEG identification with $\log_2FC$ of ± 2 , some of these terms were commonly identified in reference and *de novo*-based DEG identification. The results clearly indicate the significance and functionality of gene or gene products in carrot growth, development or genetic makeup as well as the reliability of the RNA-seq data in distinguishing and exhibiting the importance of these two contrasting cultivated carrot types. However, new GO terms or functional annotations derived from DEGs of *de novo* assembly as discussed above must be validated further to utilise the specific functional products in crop improvement program. Emphasis on these biological functions and their quantum of expression in temperate or tropical cultivars helps to explore them precisely for trait-targeted breeding in carrot improvement.

5.6.6 DEGs in cellular components (CC) in root transcriptome- *De novo* based

Among the sixteen CC terms identified for DEGs annotated in Blast2GO, cellular anatomical entity (709 genes) was among the top CC followed by intracellular anatomical structure (407) or membrane (398). Other organelles such as chloroplast, nucleus, plastid, cytoplasm *etc* were also found in the CC enrichment. In the Fisher's exact test of enrichment, genes upregulated in tropical cultivars exhibited CC terms overrepresented in 'cellular anatomical entity' and 'intrinsic & integral component of the membrane' by their enrichment in reference and test sets. Whereas, in temperate cultivar upregulated gene lists, 16 terms were related to CC in FET that showed overrepresentation of CC terms specific to photosynthetic and associated cellular structures such as plastid, chloroplast, thylakoid or photosynthetic membrane subcomponents *etc* as confirmed in reference-based annotation and GSEA as well, same results were found in temperate carrots mature root transcriptome study (Machaj *et al.*, 2018; Rong *et al.*, 2014). This clearly indicates that, the two cultivated carrots used in the present study have developed diverse anatomical structure or cellular structures during their evolution by selective forces. Wherein, cell wall or membrane related anatomical structures would have made tropical cultivars better adaptable to wider climate and 'photosynthetic cascade' is very well developed in temperate cultivar in order to accumulate higher carotenoids. Further details and evidences for CC terms in root growth and development, stress responses, reproduction, hormonal regulation in temperate and tropical cultivars is presented in next sections of this chapter.

5.6.7 DEGs in cellular components (CC) in FPL transcriptome- *De novo* based

A total of 15 subcategories were found functional in CC. Similar to root transcriptome some of the cellular processes were common in FPL tissue also like membrane, cell, intracellular *etc.* However, FPL tissue showed some unique CC terms like cellular anatomical entity, intracellular anatomical structure *etc.* These cellular processes are crucial for floral transition and formation of reproductive organs and seed formation. The GO analysis in carrot for reproductive tissue is the first of its kind; hence, these terms need to be understood in detail for further phenotypic selection during breeding for the desirable morphogenesis related genes for early flowering in temperate adapted Western cultivars in the context of development of tropical suitable Western cultivars.

5.6.8 DEGs in Molecular function for root tissue (MF)-Reference based

MF in GO analysis generally corresponds to activities that are performed by gene or gene products (*i.e.* RNA or protein) and GO terms of MF are associated with the term ‘activities’ to avoid the confusions of gene product name an MF (Gene Ontology Consortium, 2006). However, some activities are performed by molecular complexes corresponds to multiple gene products. Six GO terms were identified for MF in GO analysis whereas; seven MF terms were identified in GSEA for the tropical upregulated/temperate downregulated genes. Important MF activities found in upregulated were associated with molecular function regulator, transcription regulator, DNA-binding TF, hydrolase, transferase activity or catalytic activity *etc* in GO categorisation and GSEA.

Similarly, out of six downregulated MF in GO analysis, only one (Tetrapyrrole binding) was enriched in GSEA in temperate downregulated or tropical upregulated cultivar types. Among GO terms, DNA binding, MF regulator, transcription regulator activity, chlorophyll binding were the main MF terms.

5.6.9 DEGs in Molecular function (MF) for root tissue-De novo based

A total of 26 MF subcategories were assigned from 1598 DEGs annotated in Blast2GO where in the top functions included were associated with ‘binding’, ‘catalytic

activity’, ‘transferase activity’, ‘hydrolase activity’ ‘oxidoreductase activity’ *etc* that directly or indirectly enhances the gene or enzymatic activity in specific molecular function () or in metabolic processes ensuring normal cell activity in turn normal BP under different circumstances during crop growth. The results are in accordance with the previous results of carrot (Machaj *et al.*, 2018) Sweet potato (Firon *et al.*, 2013) and radish (Yu *et al.*, 2016). Through the Fisher’s exact test with tropical upregulated or temperate downregulated genes lists as reference set, sixteen enriched terms were found such as ‘catalytic activity’, ‘organic cyclic compound binding’, heterocyclic compound binding, ‘protein kinase activity’ *etc*. Whereas, many of the MF terms in reference set were underrepresented that the total DEG reference set in the temperate upregulated genes FET analysis. The subcategories enriched in temperate include ‘transferase, transporter activity’, ‘catalytic activity’, protein kinase activity, ‘anion binding’, ‘small molecule binding’ *etc*.

5.6.10 DEGs in Molecular function (MF) for FPL tissue-De novo based

In FPL tissue 29 MF terms were identified compared to root. A few terms were common between root and FPL tissue such as catalytic activity, nucleotide binding, hydrolase activity *etc*. However, transporter and transferase activity were some important MF terms identified in FPL tissue. During reproductive phase shift from vegetative phase, many of the flowering related genes like CONSTANTS, Flowering locus T (FT) and other will be expressed first in the leaf and will reach the shoot apical meristem (SAM) only after the long-distance transport from leaf to SAM. Hence, transporter/transferase activity could be crucial for successful transition of vegetative to reproductive phase (Kim *et al.*, 2020).

5.6.11 KEGG pathway -Reference based

A total of 24 pathways were mapped to the KEGG database in reference based DEGs where in 15 were upregulated in tropical cultivar/downregulated in temperate and 9 were upregulated in temperate cultivar or downregulated in tropical. Major and dominant pathways identified either in up or downregulated DEGs were ‘metabolic pathways’ and ‘biosynthesis of secondary metabolites’. Similar pathways were also enriched in radish taproot thickening studies with largest categories of genes assigned to these two pathways (Yu *et al.*, 2016).

Whereas, in tropical cultivars, other important KEGG mapped pathways were upregulated viz., plant-pathogen interaction, MAPK signalling pathways, phenylpropanoid biosynthesis, *etc.* The Plant-pathogen interaction, mitogen-activated protein kinase (MAPK) pathways are mostly involved in signal transduction and reported to play a crucial role in cell cycle regulation and root developmental process (Yu *et al.*, 2016; Yao *et al.*, 2013) and also importantly involved in disease response to root pathogenic nematode (Banakar *et al.*, 2020), other biotic stress (Gohar *et al.*, 2010). Hence, tropical cultivars are known for better tolerance to many of biotic stresses under varied environments and hence need to be better exploited in crop breeding of carrot. Phenyl propanoid biosynthesis is also important in releasing the phenols against the water stress and wound stress in carrot that helps in pre- and post-harvest management of carrots (Becerra-Moreno *et al.*, 2015). ‘Cyanoamino acid metabolism’ along with carbohydrate metabolism (Glycerophospho-lipid metabolism, glyoxalate and dicarboxylate, pentose and glucuronate interconversion, linoleic acid, cysteine and methionine *etc.*) is another important pathway that respond to salt stress in tomato (Zhang *et al.*, 2018), which has been upregulated in tropical carrot in the present study. The detailed information with special reference to biotic stress responses in tropical and temperate cultivar is presented in the later section in this chapter.

In case of upregulated genes in temperate carrot, other major pathways mapped to KEGG database were ‘plant hormone signal transduction’, ‘carbon metabolism’ and carbon fixation in photosynthesis organs’, ‘photosynthesis, glycolysis/gluconeogenesis’ *etc.* Plant hormone signal transduction plays a crucial role in tap root thickening (Yu *et al.*, 2016), whereas, photosynthesis plays a crucial role in formation of storage organs (Simkin *et al.*, 2020). Carbon fixation in photosynthetic organ or carbon metabolism are important in supplying methyl groups for all biological methylation reactions including nucleotide metabolism, synthesis of purines and pyrimidines and amino acid metabolism and have important role in synthesis of vitamins like Vit-B12, B6, B2 or folate and in vitamin C accumulation *etc.* (Wu *et al.*, 2017). Hence, temperate carrot are the good sources of multi-vitamins, antioxidants with nutraceutical and health benefits (Ahmad *et al.*, 2019). Growth temperatures of carrots influence the level of the sugars, carotenoids, and volatile compounds, so that growing in cool conditions results in a

higher yield and quality of carrots, while higher temperatures would increase terpene synthesis, resulting in carrots with a bitter taste. It is worthwhile to investigate the cultivation of different genotypes under various environmental conditions to increase levels of phytochemicals and enhance the nutritional value of carrot, along with the valorisation of carrot by-products.

5.6.12 KEGG pathway for root transcriptome-*De novo* based

In DEGs identified by in *De novo*-based assembly, annotation in Blast2GO elucidated 76 KEGG pathways with the top enriched pathways included ‘starch and sucrose metabolism’, glycolysis/gluconeogenesis, inositol phosphate metabolism, purine metabolism, pentose phosphate pathway, steroids, terpenoid biosynthesis, photosynthesis and metabolism of many of the vitamins (Thiamine, vitamin B6, folate, riboflavin) *etc.* Enzymes that were involved in multiple pathways of secondary metabolites, biosynthetic pathways were identified such as reductase, kinase, different transferases, decarboxylase, hexose diphosphatase *etc.*

5.6.13 KEGG pathway for FPL transcriptome-*De novo* based

A total of 40 KEGG pathways were identified in FPL transcriptome, out of which some of them were common between root and FPL tissue. Like, Carbon fixation in photosynthetic organisms, Pyruvate metabolism Glycine, serine and threonine metabolism, Flavonoid biosynthesis, Phenylalanine metabolism *etc.* were identified in FPL tissue, that are in turn important in flowering regulation in plants (Glazinska *et al.*, 2017) should further need to be studied in carrot.

5.7 Transcriptional Factors (TF) identified in root transcriptome

Transcriptional factors (TF) are proteins bind to a specific DNA sequence regulate the target gene expression, at some developmental stages or cellular processes, TF play crucial roles (Riech-mann and Ratcliffe, 2000). Our *de novo* assembled unigenes and DEGs were subjected to TF identification in the present study as detailed below.

Both unigenes (44,856) and 2239 common DEGs (1149 and 1090 up and downregulated respectively) were separately subjected to transcription factors identification by plantTF database. Total of 477 TF from Unigenes and 28 and 33 TF from up and down regulated DEGs (total 61 TF from 2239 DEGs) respectively were identified in the present study.

From the unigenes, 45 different classes (from 477 TF) of TF were identified with the highest TF included were bHLH, ERF with 40 sequences each, WRKY and C₂H₂ TF with 28 sequences each, MYB-related and bZIP with 27 each, GRAS (23), NAC (20) *etc* were most common TF. Basic helix loop helix and Ethylene responsive factor (ERF) gene are the largest families of TF involved in wide range of biological processes, such as embryo development, flower and fruit development, fruit dehiscence, cell proliferation and differentiation *etc* (Littlewood and Evan, 1995; Souer *et al.*, 1996; Roberti *et al.*, 2009; Bianchi *et al.*, 2015). ERF is a multigene superfamily of TFs that act under different growth and developmental mechanisms used by plants to respond to several biological processes and to several types of biotic and abiotic stresses. In root crops like radish, bHLH also plays a crucial role in underground storage organs (Yu *et al.*, 2016; Yao *et al.*, 2013), root growth and regulated vascular cell differentiation (Bao *et al.*, 2014). NAC TF play a functional role in plants in response to senescence and abiotic stresses and such as drought, salinity, high osmolarity, and phytohormone ABA and hence tolerate stress without major loss in yield (O'Shea, 2015). GRAS family of TF play a key regulatory function in cell division, cell differentiation, and expansion, more importantly the cambium and secondary tissue development were upregulated in Radish (Yu *et al.*, 2016) and regulates asymmetric cell division in wheat root development (Sabatini *et al.*, 2003; Drisch, and Stahl 2015).

In case of DEGs upregulated in temperate cultivar, 19 TF classes (28 total TF) were identified wherein, the top TF included were AP2, GATA, GRF were higher in number compared to tropical. *Apetala-2* (AP2) TFs is an important family involved in many biological processes and play a role in flowering, fruit growth & development especially in climacteric fruits and in ethylene signal transduction pathway (Yu *et al.*, 2016). AP2 also played crucial role in cold response and acclimatisation (Díaz-Martín *et al.*, 2005). GATA is an important light responsive TF (Teakle *et al.*, 2002) and along with AP2, it was downregulated in radish pre-cortex splitting. Growth regulatory factor

(GRF) family of TF is important in regulation of cell expansion (Kim *et al.*, 2003), known to control cell proliferation in root meristem (Challa *et al.*, 2018).

Among the tropical cultivar upregulated DEGs also 19 different TF were identified wherein, bHLH, ERF, NAC and WRKY transcription factors were higher compared to temperate. Whereas, TF that were common in both the cultivars were bZIP (2 TF each), HSF (2 TF), MYB related (one), HD-ZIP (one). As discussed as above, bHLH also plays a crucial role in flavonoid biosynthesis and helps in fruit quality (Shen *et al.*, 2014), ERF in plant development (Du *et al.*, 2013), NAC TF in plant development and stress response (Souer *et al.*, 1996) and WRKY TF is important in regulating in biotic and abiotic stresses (Eulgem *et al.*, 2000), nematode resistance (Claverie *et al.*, 2011). Hence, tropical cultivars show better tolerance to biotic and abiotic stress responses which are discussed in detailed in the later sections.

The transcriptomes (Martínez-Gómez *et al.*, 2011) offers additional molecular opportunities for the analysis of TFs linked to agronomic traits. In this regard, it is important to note the transferability of molecular information about TFs identified in the present study with other different carrot species, to evidently comprehend the synteny relationship. This synteny can result in homologous TFs from a common ancestral DNA including orthologous genes from different species or paralogous genes involving new functions (Shulaev *et al.*, 2008).

5.7.1 Transcriptional Factors (TF) identified in FPL transcriptome

Flowering is the transition from vegetative to reproductive growth and is a key developmental stage of successful seed production in flowering plants. The transition from vegetative to reproductive growth in plants is controlled by the sequential and coordinated activity of transcription factors (TFs) that integrate environmental signals, such as photoperiod, temperature and nutrient status, and endogenous cues such as gibberellin, autonomous, and plant age pathways. Over the past three decades, key TFs controlling floral transition and flower formation have been identified in *Arabidopsis thaliana* (Wils *et al.*, 2017; Kim *et al.*, 2020). TFs are crucial for the action of the RNAPs, that have been studied in the case of mRNAs and RNA polymerase II. All major processes of life depend on differential gene expression, which is generally controlled by these TFs (Kornberg, 2007; Atkins *et al.*, 2011). The transition from

vegetative to reproductive growth in plants is controlled by the sequential and coordinated activity of transcription factors (TFs) that integrate environmental signals, such as photoperiod, temperature and nutrient status, and endogenous cues such as plant age. Floral homeotic transcription factors (TFs) act in a combinatorial manner to specify the organ identities in the flower (Chen *et al.*, 2018).

A total of 539 sequences were annotated to TF in FPL tissue transcriptome, and these have been classified into 53 TF families. TF families expressed in the FPL tissue transcriptome were high compared to the matured storage root transcriptome of the carrot. Among common 1671 DEGs of FPL tissue, 29 loci encoded for TF with 15 different TF families. Of these 29 TFs differentially expressed, 8 TF were induced upregulated in temperate or downregulated in tropical cultivar. While, 21 TF were downregulated in temperate, in other words upregulated in tropical cultivars. This suggests the possible major transcriptional reprogramming occurred to manifest these two cultivars phenotype and flowering pathway.

In TF families in FPL tissue unigenes, bHLH (44) and ERF (37) and C₂H₂ (34) TF family were highly enriched. There were 34 genes that code for C₂H₂ TFs were induced in FPL tissue transcriptome followed by MYB (28), HD-ZIP (21) bZIP(29), C3H (21), MYB_related (21), NAC (21), WRKY (18), GATA (16), Trihelix (16), TALE (14) were other major TF identified. Among other TF, AP2 (8), GRAS (21), B3(14), G2-like (12), Dof (11), TCP (8), E2F/DP (7), MIKC_MADS (11), M-type_MADS (4), LFY (1) and SBP (9) were also identified in unigenes sequences.

Whereas, TF identified in common 1671 DEGs, 21 out of 29 were upregulated in tropical or downregulated in temperate. The temperate downregulated DEGs were as follows. YABBY (2), MIKC-MADS (5), HD-ZIP (3), MYB-Related (5), MYB (1), bZIP (1), TALE (1), B3(1) and NF-YA (1), TCP (1) were downregulated in temperate. On the other hand, the upregulated TF in temperate cultivars were, bZIP (3), WRKY (1), ERF (1), C₂H₂ (2) and MYB (1) were found as upregulated.

C₂H₂ abundant in our FPL transcriptome is known to activate the flowering repressor gene FLC (Zhoa *et al.*, 2019). TCP, ZFP, MYB, WRKY, and bHLH families may be involved in the transcriptional regulation of flowering genes, MADS-box TFs are the key regulators of flowering (Zhoa *et al.*, 2020). Several MIKC type MADS box

genes are known to play a crucial role in flowering time control like FLC, SOC1 that promote flowering under favourable condition (Smarrh *et al*, 2019).

5.8 Genetic variant analysis

Identification of genetic variants from RNA-seq is the most crucial and valuable information for plant breeders to utilize them in breeding programs such as QTL mapping, cultivar identification for target specific traits, marker-assisted breeding *etc.* Hence, RNA-seq data was subjected to genetic variant analysis such as microsatellites identification (from unigenes identified in *de novo* assembly) and SNP and *indel* markers identification (from reference-based analysis). The details of genetic variants identified in root transcriptome and FPL transcriptome are discussed below.

5.8.1 Identification of Single nucleotide polymorphism (SNP) and *Indels* in storage root transcriptome

A total of 236071 SNPs were identified of which 134335 (56.90 %) transitions (Ts) and 101736 (43.10 %) transversions (Tv) and 1.32 Ts/Tv ratio was identified for root transcriptome data in the present study from six paired -end RNA-seq libraries. In the previous *de novo* assembly and identification of SNPs in carrot by Iorizzo *et al*, 2011, 20,058 SNPs in 7,684 contigs were identified of which 13,756 (69%) were transitions (A/G, C/T) and 6,302 (31%) were transversions (G/T, C/G, A/T, A/C). SNPs identified from the present RNA-seq data of the present study is highly enriched with SNPs and the present analysis is in comparison with the reference data (considering phred score (Q) at least 20 and depth of 10) as compared to previous report based on *de novo* assembly prior to the publication of carrot genome. Hence, SNPs identified in the present study are most useful genomic information for carrot breeders especially in the breeding program that involves temperate and tropical cultivars. Among the transitions, T>C and A>G (~17.5 %) are more frequent than C>T and G>A conversions (10.87 %). Whereas, in transversions, C>A and G>T (~6 %) than other type of transversions.

Among the total of 20866 indels identified from six root RNA-seq libraries, single base deletions (-1) were 45% (9505) followed by single base addition (1bp) with 40.45 % (8435). Only 4.99% of the indels were belonging to addition of two bases (2bp). As per our knowledge this is the first report on identification of indels from

whole transcriptome data in carrot. Again, this type of genetic variants is very useful in plant breeding activities from gene tagging/QTL identification and marker assisted breeding program.

Further, it is necessary to identify polymorphic SNPs and indels between tropical and temperate carrot cultivars and annotation is necessary for their validation and utilization in the breeding program.

5.8.2a Microsatellite identification in storage root transcriptome

For microsatellites identification, unigenes containing 44856 contigs (longest isoforms) were used in microsatellite identification tool (MISA) and identified 4437 putative perfect microsatellites out of which 520 were compound SSRs. In the previous studies on microsatellite identification in carrot (Iorizzo *et al.*, 2011), 8823 putative SSRs in 6,995 contigs were identified of which 323 were compound SSRs and remaining perfect SSRs.

In our analysis, following criteria were applied for SSR identification viz., minimum repeats for each perfect SSR type were set to 12 for Mono-, 7 for Di, 5 for Tri, 4 for Tetra, 4 for Penta, 4 for Hexanucleotide and motif standardization level was set to Level 3. Highest perfect SSRs were dinucleotide repeats (45.44%) followed by mono nucleotides (27.92%) and the least were pentanucleotide repeats (1.26%). Among dinucleotides, AG type repeat (1255) were found to be highest followed by AC type repeats. Among tri nucleotide repeats AAG (179/873) was highest followed by ATC (124/873). Among tetranucleotide repeats, ATC (48/180) was the most abundant followed by AAAT (32/180).

5.8.2b Microsatellite identification in FPL tissue transcriptome

Based on the motif repetition, these microsatellites were further categorized into mononucleotide (196), followed by di-(6365), tri- (750), tetra-(152), Penta-(33) and hexanucleotides motifs (381) (Table 27). AG dinucleotide motif was the dominant for dinucleotide repeats followed by AA. while AAG was most abundant for trinucleotide microsatellites. Among tetranucleotide repeats, ATAC was the most abundant followed by AAAT.

Uncu and Uncu, 2020 mined SSR markers by in silico analysis from carrot genome assembly and identified 55,386 markers, of these 51,160 SSRs were single-locus markers and 4,226 amplified more than one locus and 578 markers showed low copy transferability with Fennel. In their study also dinucleotides showed highest frequency (54.51 %) followed by mononucleotide (39.63%).

Further analysis of polymorphic microsatellites and cultivars specific SSRs need to be performed in order to utilize them in the carrot improvement program to select the cultivar specific segregants in the segregating populations for the trait- targeted breeding. Identification of molecular markers have numerous uses in plant breeding, including linkage map development, quantitative trait loci mapping, marker-assisted selection, parentage analysis, cultivar fingerprinting, genetic diversity studies, gene flow and evolutionary studies (Zalapa *et al.*, 2012). Recently, transcriptome sequencing has become a viable method for microsatellite mining and SSR marker development, and has been utilized in many species (Zhong *et al.*, 2021).

5.8.3 Identification of Single nucleotide polymorphism (SNP) and Indels in FPL tissue transcriptome

A total of 445144 SNPs were identified of which 245786 (55.21%) transitions (Ts) and 199358 (44.78%) transversions (Tv) and 1.32 Ts/Tv ratio was identified for FPL transcriptome data in the present study from six paired -end RNA-seq libraries. Among the transitions, A>G and T>C were more frequent than C>T and G>A conversions and *etc.* Where as in transversions, G>A and C>T was more common than other type of transversions. Total of 25200 indels were identified among the tropical and temperate carrot FPL transcriptome.

5.9 Gene regulatory network (GRN) analysis for storage root development

It is well-known fact that, vascular cambium (VC) is the master regulator in deciding the xylem and phloem thickening during secondary storage root formation in

almost all the crops where root is the edible portion. Continuous cell division, more specifically anticlinal type of cell division in VC provides precursor cells for xylem secondary growth inside the VC and phloem extension outside the VC cells by their radial growth of roots (Hoang *et al.*, 2020a). In this process number of stress responsive TF, vascular cambium partitioning genes/TF and carbon partitioning related genes play greater role by interconnecting network was explained in detail by Hoang *et al.* (2020b). With the meta-analysis of genes involved in VC partitioning from *Arabidopsis* and their orthology search with other root crops such as radish, beetroot, carrot that are important secondary storage roots where VC clearly partitions xylem and phloem tissues.

Carrot storage roots develop via thickening of the primary (tap) root. Storage root is formed by the thickening of the hypocotyl and upper root, which results from the activity of the vascular cambium (Zaki *et al.*, 2012; Hoang *et al.*, 2020a). Hence, considering the reference genes involved in storage root formation and VC development of *Arabidopsis*, we performed GRN analysis to understand the genes responsible for storage root formation and their interaction with stress responsive genes in temperate and tropical cultivars carrot separately.

5.9a GRN for storage root development and Photosynthesis regulation in temperate cultivar

In case of temperate carrot cultivar, as discussed earlier, photosynthetic cascades and photosynthesis related genes are upregulated in GO in BP or CC subcategories. For the secondary storage root development that grows in lateral direction, photosynthates efficiently translocate from sources (leaf) to these storage roots through sieve elements, companion cells and plasmodesmata. Wherein, sucrose (by product of photosynthesis) converted to other sugar compounds and metabolites through the process of starch, sucrose metabolism (Hoang *et al.*, 2020a). Hence, in the GRN clustered genes which are interconnected each other are mostly involved in photosynthetic related activities (ex: LSH10, PSAH, PSAN, PSAD, LHCB, PSB, *etc.*), stress responsive genes like WRKY70, ERF-71, MKK9, MYB15, CAMBP25 *etc.* and with root vascular developmental genes (TRN2, LSH10, ATG13610, AT1G56150, SP1L2, HB6, AT1G66830). TRN2 is involved in the basipetal transport of auxin (IAA) that modulates growth and organs organization, as well as cell differentiation. TRN2

regulates shoot apical meristem (SAM) organization in the peripheral zone and required for initial meristematic divisions in the epidermal/lateral root cap leading to the formation of epidermal cells and a clone of lateral root cap cells, as well as for the maintenance of the radial pattern of cell specification in the root (Lee *et al.*, 2019). Thus, this regulates the distinction between the lateral root cap and epidermis. LSH10 (developmental regulator by promoting cell growth in response to light) (Zhao *et al.*, 2004), ATG13610 (Expressed hypocotyl root), AT1G56150 (plays a role in the regulation of cell expansion, root meristem patterning and auxin transport), SP1L2 Mutant plants exhibit altered patterns of root and organ growth as a result of defective anisotropic cell expansion (Nakajima *et al.*, 2006). HB6 involved in the negative regulation of the ABA signalling pathway (Carianopol *et al.*, 2020), AT1G66830 (Leucine-rich repeat protein kinase family protein interaction with Phloem intercalated with XYLEM (PXY)/TDIF RECEPTOR PROMOTES cell divisions in the cambium, inhibits xylem differentiation and controls vascular patterning *etc* (Hoang *et al.*, 2020a). In the temperate cultivar, the cambium layer does not look distinct due to deeply pigmented carotenoids in both xylem and phloem tissues (self-core). Hence, it is hypothesized that the xylem differentiation and vascular patterning is not observed in GRN in this cultivar.

5.9b GRN for storage root and stress response in tropical cultivar

GRN analysis in tropical carrot cultivar, in contrast, showed highest number of stress responsive genes/TF (LOX2, LOX3, LOX4, WRKY46, MYB15, ERF4, PUB24, KIN3, NAC036, NLA *etc*) along with few roots growth-related genes (PERK10, AT2G26380 Leucine-rich repeat (LRR), LSH6, HCA2 *etc*) than photosynthesis network genes (0) or carbohydrate metabolism (NANA). LOX2, LOX3, LOX4 involved in plant growth and development, pest resistance, and senescence or responses to wounding (Grebner *et al.*, 2013).

LOX2 was shown to contribute the majority of jasmonate synthesis upon wounding and osmotic stress and during senescence in leaves (Bell *et al.*, 1995; Glauser *et al.*, 2009). LOX3 and LOX4 are required for the development of fertile flowers (Caldelari *et al.*, 2011). LOX6 shows overall low expression (Bannenberg *et al.*, 2009). Recently, it was reported that LOX6 contributes to the fast accumulation of JA and JA-

Ile in wounded leaves and is required for the fast increase of JA and JA-Ile in distal leaves after wounding (Chauvin *et al.*, 2013). In contrast to leaves and flowers, little is known on jasmonate biosynthesis and function in roots. Mutant plants defective in LOX6 are impaired in stress-induced jasmonate synthesis and are more susceptible to drought and detritivore feeding (Grebner *et al.*, 2013).

WRKY46 transcription factor involved in the regulation of osmotic stress responses and stomatal movement. MYB15 involved in drought and salt tolerance, may enhance expression levels of genes involved in abscisic acid (ABA) biosynthesis and signalling, as well as those encoding stress-protective proteins (Ding *et al.*, 2009). ERF4 antagonizes JA inhibition of root elongation (McGrath *et al.*, 2005). PUB24 E3 ubiquitin-protein ligase PUB23 negatively regulates water stress response, act as negative regulator of the immunity triggered by the pathogen-associated molecular patterns (PAMPs) Trujillo *et al.*, 2008. KIN3 a putative serine/threonine-specific protein kinase have defence response against pathogens and viruses (Romeis, 2001). NAC036 identified as master regulator of xylem vessel differentiation and secondary cell wall thickening (Nakano *et al.*, 2015). NLA036 mediates E2-dependent protein ubiquitination, plays a role in salicylic acid-mediated negative feedback regulation of salicylic acid (SA) accumulation and involved in the overall regulation of SA, benzoic acid and phenylpropanoid biosynthesis (Hoang *et al.*, 2020a).

Among root growth related genes like PERK10 is a proline-rich receptor-like protein kinase could be involved in the negative regulation of root growth (Hwang *et al.*, 2016), AT2G26380 Leucine-rich repeat (LRR) is expressed in hypocotyl and decides xylem differentiation upon interaction with XYLEM (PXY)/TDIF RECEPTOR PROMOTES Chakraborty *et al.*, 2019. Whereas, LSH6 is a probable transcription regulator that acts as a developmental regulator by promoting cell growth in response to light (Li *et al.*, 2015). HCA2 is a Dof-type zinc finger DNA-binding family protein; induces the formation of interfascicular cambium and regulates vascular tissue development (Hoang *et al.*, 2020a).

Photosynthesis was down regulated in tropical carrots as compare to temperate cultivars in both reference-based assembly and *de novo*-based assembly annotation as well as GO enrichment. While, in GRN analysis tropical carrots showed abundance for

stress responsive genes for both biotic/abiotic stresses than photosynthesis or carbon metabolism. NANA an aspartic proteinase that regulates endogenous sugar levels (e.g., sucrose, glucose and fructose) by modulating starch accumulation and remobilization was detected in the complex network Paparelli *et al.*, 2012. NANA is also involved in the maintenance of the shoot apical meristem (SAM) influences general morphology and development (Albert *et al.*, 2014).

Here, storage root formation in temperate carrot, reduction in carbon flow towards phenyl propanoid biosynthesis pathway and its delivery into carbohydrate metabolism and starch biosynthesis are recognised as major events involved in storage root initiation and development. Not surprisingly, among the DEGs enriched in temperate carrot most were related to photosynthesis, carbon metabolism, carotenoid related genes were all upregulated in temperate carrot in contrast to tropical carrot. However, tropical carrots are vastly enriched with number of biotic like bacteria, fungus, virus, insect and abiotic stress responsive genes (drought, high temperature, cold acclimation, response to wound and salt stress *etc*). Comprehensive understanding of pathways is necessary for utilization in carrot breeding programs.

5.10 Candidate genes related to root adaptability and flowering identified in the present study

Adaptability could be defined as the ability of a crop or variety to positively respond to changes in agricultural conditions. In this regard, root plasticity is considered as a target for the development of more productive crops under varied environmental conditions as phenotypic plasticity has many potential ecological, and physiological benefits (Schneider and Lynch, 2020). Carrot crop that has been classified into temperate adapted Western carrot and tropical adapted Asiatic carrot have an ancestral relationship with each other with recent domestication history (Ellison *et al.*, 2018). Despite being wider adaptation to warmer climates, tropical carrots have been least explored in varietal improvement due to the lack of knowledge of the tropical adapted carrot at the molecular level. Comparative transcriptome profiling of these wider

adaptable temperate carrots with cool climate adaptable temperate type provides broad spectrum and dynamic range of expression level among these for better understanding of the candidate genes that are differentially expressed during the critical stages of lifecycle in a single experiment without creating multiple stresses. Hence, in the present study after sequencing two important tissues, at two developmental stages between two contrasting cultivars, putative candidate DEGs that are involved in root development, physiological fitness in terms of photosynthesis efficiency, source-sink translocation like carbohydrate metabolism, innate immune response highlighting the constitutive defense response to biotic and abiotic stresses and also their ability to induce flowering and set seeds under wide climatic conditions were identified from a high quality transcriptome assembly constructed in the present study with significant DEGs with minimum Log₂ fold change ± 2 at low P-value of 0.01 and false discovery rate <0.05.

A) Root development

Being underground and often ignored, root is referred to “hidden half” of the plant. A strong and well-established root system plays an important role in plant performance and yield in case of crop plants (Lynch 2011; Paez-Garcia *et al.*, 2015). The variation in RSA is very prominent between tropical and temperate carrots. Tropical carrots have clear xylem phloem patterning (GO:0010051), where as temperate carrots have deep uniform pigmentation without visible xylem phloem patterning. Which was found to be regulated by xylem and phloem pattern formation (GO:0010051-02), cotyledon vascular tissue pattern formation (GO:0010588-1), Stem vascular tissue pattern formation (GO:0010222-1) genes in tropical plants. Vascular patterning related DEGs are absent in temperate-type carrots.

Apical meristems are found at the apex, or tip, of roots and buds, allowing roots and stems to grow in length and leaves and flowers to differentiate. Roots and stems grow in length because the meristem adds tissue “behind” it, constantly propelling itself further into the ground (for roots) or air (for stems). Apical meristems differentiate into the three basic types of meristem tissue which correspond to the three types of tissue: protoderm produces new epidermis, ground meristem produces ground tissue, and

procambium produces new xylem and phloem. These three types of meristems are considered primary meristem because they allow growth in length or height, which is known as primary growth. Secondary meristems allow growth in diameter (secondary growth) in storage root plants. Secondary growth adds to the diameter of a stem or root; vascular cambium adds xylem (inward) and phloem (outward) (Zaki *et al.*, 2012; Lee *et al.*, 2019; Haung *et al.*, 2020a). Hence, difference in vascular cambium related DEGs expression and gene counts are probable reasons for clear xylem-phloem patterning observed in tropical carrots.

Lateral root formation in tropical is govern by two genes (GO:0010311) and in temperate types it is govern by four genes (GO:0048527-2; GO:0010102-2).

B) Carbohydrate and lipid metabolism:

Assimilation of carbohydrate in storage roots is important part of secondary storage root development. In the storage root of carrot, 1.2–6.44% of the mass is dietary fiber, and 80.94% of the dietary fibre is cellulose Carrot is well known as being a good carotenoid provider. Moreover, carrot roots also contain many other beneficial contents, including vitamins, carbohydrates and minerals. According to Li *et al.* (2002) sugars, glucose, fructose and starch are the main types of carbohydrates in carrot storage roots. Hence list of genes involved in carbohydrate metabolism in tropical and temperate carrots would help to understand assimilation of carbon in storage root. The relevance of root secondary growth for CO₂ sequestration is further highlighted in the meta-analysis by Poirier *et al.* (2018) where root suberin content was identified as one of the most influential promoters of soil organic matter stabilization. Greater research in understanding the potential that root secondary growth has for carbon storage below-ground will only become more critical as climate change progresses.

Proportion of genes involved in polyacetylenes (lipid) metabolism is high in tropical carrots than temperate carrots. GO terms like, Monoterpene biosynthetic process (GO:0043693), Sesquiterpene biosynthetic process (GO:0051762), Carotenoid biosynthetic process (GO:0016117), Sterol biosynthetic process (GO:0016126),

Xanthophyll biosynthetic process (GO:0016123), Regulation of membrane lipid metabolic process (GO:1905038), Unsaturated fatty acid biosynthetic process (GO:0006636), Phospholipid binding (GO:0005543) and, Phosphatidylinositol-3-phosphate biosynthetic process (GO:0036092) known for lipid metabolism were found abundantly in tropical carrots, however, completely absent in temperate carrots. Polyacetylenes are a large group of non-volatile lipid compounds that comprise at least two, usually conjugated, triple carbon–carbon bonds. They are primarily produced by members of the *Apiaceae* and *Araliaceae* families, with carrot being the major dietary source of polyacetylenes.

Some polyacetylenes, such as falcarinol-type polyacetylenes, are bioactive phytochemicals with proven antifungal, antibacterial, allergenic, anti-inflammatory, anti-platelet aggregatory and in vivo and in vitro anticancer properties (Cavagnaro, 2019). In addition, these compounds are largely responsible for the bitter unpleasant taste perceived in some carrots. Hence, understanding the biosynthesis pathways of these polyacetylenes would help to exploit tropical carrots for drug industries.

C) Photosynthesis:

Nine photosynthetic efficiency related terms were identified for temperate cultivars. viz., Photosynthesis (GO:0015979), Photosystem I assembly (GO:0048564), Photosynthetic electron transport chain (GO:0009767), Photosynthesis, light harvesting in photosystem I (GO:0009768), Photosystem II assembly (GO:0010207), Photosynthesis, light harvesting in photosystem II (GO:0009769), Photosystem II stabilization (GO:0042549), Regulation of photosynthesis (GO:0010109), Photosystem II repair (GO:0010206) regulated by total of 15 gene counts. Tropical cultivars had only two photosynthesis related terms. Photosynthesis (GO: 0015979), Regulation of photosynthesis (GO:0010109) govern by only two gene counts.

Differential expression level of genes involved in post-phloem transport and the relevant activity of their encoded enzymes further need to be studied for supporting symplastic unloading in temperate carrots storage roots. Plasmodesmata provide

symplastic continuity between most adjacent cells in higher plants and thus substantially contribute to phloem unloading into sink tissues (Oparka and Cruz, 2000.) Hence, its necessary to study the plasmodesmata number and structure between temperate and tropical carrot plants. Patrick and Offler (1996) found that predominantly symplastic unloading was associated with a high transport capacity and low resistance. This is in support of low biotic and abiotic stress responsive genes identified in temperate carrots with high photosynthetic efficiency identified in GRN developed for temperate carrot.

D) Candidate genes for constitutive defense response for biotic factors

As tropical carrots are rich reservoirs for biotic and abiotic stress related breeding programs in carrot. Tropical cultivars are having highest classes (Nineteen) of biotic stress responsive genes against important pathogens like fungi (14), bacteria (10), nematode (4), virus (2), insect (1) with systemic acquired resistance (3), innate immune response (1) and various other important mechanism like Defence response by callose deposition in cell wall (1) by which Pre-invasion defense mechanisms can be exploited, especially the one that is conferred by preformed or inducible physical barriers such as callose and antimicrobial compound deposition at the site of attempted penetration (Cao *et al.*, 2011), Then, Hydrogen peroxide-mediated programmed cell death (1) wherein, Hydrogen peroxide (H_2O_2) has been shown to mediate activation of the three major and well-studied *Arabidopsis* MAPKs, MAPK3, MAPK4 and MAPK6, through MAP3Ks and other kinases (Rentel *et al.*, 2004; Teige *et al.*, 2004; Tan *et al.*, 2013) and other against different biotic stresses. Whereas, temperate carrot storage root was poorly enriched with different biotic stress responsive classes (four), bacterium (4), fungus (1) Salicylic acid biosynthetic process (01), Response to Jasmonic acid (02). Signifying the importance of utilizing tropical carrots in biotic stress related breeding programs.

E) Candidate genes for constitutive defense response for abiotic factors

Tropical carrots were enriched with highest abiotic stress response classes (22) govern by 72 different genes; whereas temperate carrots had only 14 different classes govern by almost half number of genes (33) against various abiotic stresses. Tropical

carrots had highest genes associated with Response to cold (12), followed by Response to water deprivation (11) then Response to salt stress (9). Temperate carrots had only 14 different classes with total of hardly 33 genes against various abiotic stresses. Temperate carrots had more genes associated with Response to cold (7), followed by Response to light stimulus (6) then Response to salt stress and Response to high light intensity (3). When we compare the number of genes across all the classes of abiotic stresses tropical carrots had highest gene counts than temperate carrots. This represents the importance and wider adaptability of tropical types in various climatic challenges than temperate carrots.

High soil salinity is considered a major threat for agricultural productivity in semiarid or coastal areas due to its increased occurrence in irrigated lands, which account for a major part of world food production (Flowers 2004). Thus, salinity is also an important abiotic stress limiting carrot crop cultivation. Salt stress persistence can be increased with tropical carrots utilization worldwide. Constitutive activation of AtMAPK4 and AtMAPK6 rendered plants more tolerant to cold and salt stress (Teige *et al.*, 2004).

F) Diverse stresses, unique adaptive responses and tolerance mechanisms in tropical and temperate carrots

Abiotic stress factors such as drought, cold, heat and salt stress share similar properties such as the imposition of secondary oxidative stress but these also possess unique features. Calcium signature responses during early stress perception reliant on the type of stress (Whalley and Knight 2013). Gene expression profiles are remarkably different under different stress factors (Rabbani *et al.*, 2003), and exceptional and often contrasting physiological and biochemical responses are detected, such as the contrasting plasma membrane lipid remodelling during heat and cold stress (Ruelland and Zachowski 2010) and opposing stomata responses during drought and heat stress (Zhao *et al.*, 2013). Stomatal opening for example is beneficial under heat stress, however it might increase the pathogenicity of stomatal invading pathogens, while drought induced stomatal closure might decrease pathogenicity. Heat stress in many

occasions' results in the breakdown of R-gene mediated resistance (Marques de Carvalho *et al.*, 2015; Zhu *et al.*, 2010).

Hence, the present work effort led to the identification of genes involved in the secondary root growth across temperate and tropical root crops, which are highly connected in terms of their expression patterns and functions. Future engineering schemes for the improvement of root crop yields via the possible manipulation of key nodes in GRNs of tropical carrot photosynthesis or carbon metabolism pathway would help to increase yield and customer need of tropical region and may increase global production rate of carrot types across different ecological conditions.

G) Flowering and reproductive biology of temperate and tropical carrots

Present knowledge on carrot reproductive biology is limited if compared to the model plants and inadequate, despite powerful genetic research on carrot and their wild relatives over the last decades. A possible reason for this is the prioritization of research on economically relevant traits such as yield, quality, or resistance (Linke *et al.*, 2019). A focus on reproductive biology is missing but important for breeders and seed producers. The carrot germplasm is a valuable resource for plant-breeding efforts including improvement of crop yield and protecting against pests and diseases. Therefore, understanding the genetic composition and reproductive biology within the cultivated germplasm (temperate and tropical types) has been an important goal of carrot researchers. The current knowledge on transcriptomics of floral primordia including young leaf (**FPL-tissue**) at **vegetative to reproduction transition phase** of tropical cultivars regarded as annual types with early flowering habitat and temperate types also known as biennial types with late flowering habitat helped to identify candidate genes involved in flowering. Further the obtained information helps to introduce characters from late-flowering cultivars to early flowering germplasm in breeding programs with more accuracy.

Since leaf is the primary receptor of light and the genes related to flower induction expressed in the leaf will have long distance transport to shoot apical

meristem (SAM), an important tissue for vegetative to reproductive transition, where these floral inducing genes will be expressed (Tsuji, 2017). Synchronization of many developmental processes includes communication among different locations in the plant, one of the simplest concepts being the need to balance root and shoot growth. In relation to long-distance signalling in flowering time control, the preferred photoperiod sensing organ (leaf) is spatially separated from the site of response (shoot apical meristem, SAM). The sensing of vernalization has more options. Although classic vernalization models propose that cold perception occurs locally in the meristem that itself switches into floral development (Dennis *et al.*, 1996; Amasino, 2010), other parts of the plant are capable of sensing cold and acting to regulate flowering time at a distance (Reid and Murfet, 1975; Searle *et al.*, 2006).

However, floral primordia are the by-product of an irreversible transition from vegetative to reproductive phase and the genes expressed during this phase in the floral primordia between two contrasting cultivars would provide the clear evidence of the pathway that is operating or dominating in this process. Hence, both leaf and floral primordia were selected for RNA-seq from tropical and temperate carrot cultivars. The data was subjected to *de novo* assembly and DEGs were identified in FPL tissue transcriptome following the same procedure as explained in root transcriptome. The results are discussed here with supporting the GO terms identified for tropical and temperate carrots floral transition from vegetative to reproductive phase and seed development related DEGs clustered to various seed developmental stages identified from FPL tissue transcriptome. In which, set of target genes identified for floral transition, pollen development, overall flower morphogenesis and seed development were highest in tropical carrots (46) than temperate carrots (3). This shows the diverse flowering response and different pathways between two cultivars. Clear understanding of the pathways can be utilized efficiently in breeding programs for introducing biotic or abiotic stress related genes from early flowering tropical types into late flowering temperate types and also introducing important domesticated traits (highest carotenoid, more uniform, high photosynthetic efficiency *etc.*) from late flowering temperate carrots into tropical types. The detailed list of GO terms related to reproduction was extracted

from the annotation file for tropical upregulate and temperate upregulated genes separately.

Flowering is the transition from vegetative to reproductive growth and is a key developmental stage of successful seed production in flowering plants (Bäurle and Dean 2006). To assure timely flowering in various environments, plants have formed several main genetic mechanisms that are regulated by external and internal signals. Plants utilize photoperiod signals to perceive a seasonal change. Photoperiod (or day length) regulates flowering time in many plants. The core components involved in the photoperiod pathway, such as CONSTANS (CO) and FLOWERING LOCUS T (FT), are well conserved among many plant species (Kobayashi and Weigel 2007). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis* (An *et al.*, 2004). In *Arabidopsis*, the level of FT mRNA expression determines the timing of bolting (Turck *et al.*, 2008; Corbesier *et al.*, 2007). As a facultative long-day (LD) plant, *Arabidopsis* exhibits accelerated expression of FT under LD conditions, but lowered expression under short-day (SD) conditions. Thus, the transcriptional activation of FT is critical for the induction of floral transition in the photoperiod pathway (Searle and Coupland 2004).

Carrot (*Daucus carota* L.) is a biennial plant requiring vernalization to induce flowering, but long days can promote its premature bolting and flowering. The basic genetic network controlling the flowering time has been constructed for carrot, but there is limited information on the molecular mechanisms underlying the photoperiodic flowering response. The published carrot genome could provide an effective tool for systematically retrieving the key integrator genes of GIGANTEA (GI), CONSTANS-LIKE (COL), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and FLOWERING LOCUS T (FT) homologues flowering pathway. However, other important genes like FLC, SVP, GI, PHYA, PHYB, GAI, AP2, LEAFY, FKF1, MIR172, CCA1 and TOC1 were obtained by orthology search *Arabidopsis* genes against carrot genome by Blastx. Further based on homology, primers were designed to each of the genes using primer3 for qPCR. It was ensured that the primers show perfect

complementation with our *de novo* assembled transcriptome. *Arabidopsis thaliana* is also a biennial, long day plant with obligate vernalization requirement like carrot. Over the past three decades, key genes controlling floral transition and flower formation have been well studied in *Arabidopsis thaliana* (Andrés *et al.*, 2012; Pajoro *et al.*, 2014; Wils *et al.*, 2017; Kim *et al.*, 2020). Genes like GIGANTEA (GI), CONSTANS-LIKE (COL), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and FLOWERING LOCUS T (FT) were identified in FPL tissue RNA-seq were subjected for qRT-PCR validation to ensure the reliability of results and expression pattern.

GRN analysis for flower transition between temperate and tropical carrots:

GRN clearly showed the interaction between COL, COL1, COL2 with FKF, DOF, GI and GAI mediated by circadian gene TOC1 and SPL4 and SPL5, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (SPL) floral activators, which induce the expression of FT by targeting repressor *miRNA156*. Expression of SPL upregulates the level of *miRNA172* and induce flowering in carrot.

5.11 Validation of RNA-Seq Results with Quantitative Real-Time PCR (qRT-PCR)

5.11a Root transcriptome validation:

The fold change values from RNA-seq and qRT-PCR for eight of the nine genes (G2 to G9) genes exhibited concordant tendency of gene expression. Eight genes were in agreement with the direction of the change fold. G1 had discordant tendency of expression between the methods of RNA-seq and qPCR expression in opposing directions. As eight genes are concordant and one gene is non-concordant, obtaining concordant results for a higher number of random selection of genes with good correlation between results obtained with qPCR and with RNA-seq.

According to current annotation of root transcriptome for these genes, G2 (ubiquitin-conjugating enzyme E2 20-like; LOC108208449), G3 (polyubiquitin; LOC108227972), G4 (probable WRKY transcription factor53; LOC108209169) are

involved in root development. Then, G5 (probable glycosyltransferase At5g03795; LOC108193533) involved in carbon metabolism and G6 (cysteine proteinase inhibitor B; LOC108211283) is involved in seed germination. Whereas, G7 (ATPase WRNIP1; LOC108207021), G8 (DNA repair protein XRCC2 homolog; LOC108216678), G9 (histone deacetylase 6; LOC108200546) are involved in innate immune response (Plant pathogen interaction). While, G1 (mechanosensitive ion channel protein 6-like; LOC108226635) function in homeostasis and signalling pathways, their interactions with other genes related to growth and development, responsive to light, hormone, and stress. Further, comparison of selected genes expression pattern between temperate and tropical carrots by qPCR showed that ubiquitin-conjugating enzyme E2 20-like and DNA repair protein XRCC2 homolog were the only DEGs found to be up-regulated in temperate carrots. While, remaining genes like probable WRKY transcription factor53 are involved in root development, probable glycosyltransferase involved in carbon metabolism, ATPase WRNIP1 involved in innate immune response (Plant pathogen interaction), mechanosensitive ion channel protein 6-like; LOC108226635) function in homeostasis and signalling pathways are found to be up-regulated in tropical carrots

5.11b FPL-tissue transcriptome validation:

Comparing the qPCR results with the sequencing results of FPL tissue, the expression trend in the five differentially expressed genes of that are related to flowering viz., GI (GIGANTIA), CO (CONSTANS); PhyA (PHYTOCHROME A); AP2 (APETALLA2); FKF1: (FLAVIN BINDING KELCH REPEAT, F-BOX-1) between tropical and temperate genotypes was consistent with that of the sequencing results, indicating the reliability of the sequencing method. Hence, RNA sequencing (RNA-seq) results are more accurate as this technique provide information about gene expression at the genome-scale level (*i.e.*, quantify the expression level of all genes). While, qRT approaches can be used to measure the expression levels of a smaller subset of genes. RNA-seq are expected to be reliable as they are based on a sufficient number of biological replicates(three), the added value of validating them with qPCR for five genes increase the accuracy of RNA-seq data. RNA-seq methods and data analysis approaches are robust enough to not always require validation by qPCR and/or other

approaches, although there are situations where this may be of added value (Coenye, 2021).

Expression pattern of flowering related genes:

GI can directly bind and activate FT expression, by out-competing the binding of FT repressors to the promoter of FT during the daytime. GI expression was high for tropical plants and it was found to be low for temperate carrots because the leaf samples were collected for qRT validation only after floral transition in plants. Immediately after floral transition, GI expression will be reduced same as if there was no flowering induction (Kim *et al.*, 2020).

GI has been shown to enhance the expression of miRNA172, which can subsequently cause a reduction in the AP2 repressor gene transcripts, thus resulting in a switch from transcriptional repression to activation of FT. The *A. alpina* ortholog of *Arabidopsis* AP2, PERPETUAL FLOWERING 2 (PEP2), enhances the expression of PEP1 in the vernalization pathway, thus blocking floral transition. Hence, in the present work AP2 expression is down in all samples collected after floral transition between tropical and temperate carrots. Expression of FT, dcSOC1 is high among the same samples. miRNA172 expression also gets down regulated along with GI after floral transition signal is perceived by plants (Kim *et al.*, 2020). These results clearly show that the flowering inducer genes expression gets down immediately after the signal is perceived by plants (Kim *et al.*, 2020). SOC1 integrates multiple flowering signals to interact with multiple MADS-box proteins and regulate the expression of flowering genes across all experimental conditions (Kim *et al.*, 2020).

5.12 Assessment of stability, adaptability and yield performance of carrot (*Daucus carota sub spp. sativus*) cultivars

The success of crop improvement and production activities can be enhanced with scientific information generated from genotype-environment interactions. GEI reduces the association between phenotype and genotype which result in relative ranking and stability differences of genotypes across environments (Spaldon *et al.*,

2017; Raj *et al.*, 2019; Kumar *et al.*, 2019; Hugo Ferney *et al.*, 2006). This study was conducted with the objective to identify stable and adaptable carrot genotypes under various environments of Northern dry zone of Karnataka. Considering the performance, regression coefficient and deviation from regression, the following conclusions were made for the stability analysis conducted in the present study.

Eberhart and Russell (1966) model considers both linear (b_i) and non-linear (S^2_{di}) components of genotypes $\times$ environment interaction for predicting the performance of a genotype, an ideally stable/adaptable genotypes would be the one having unit regression coefficient ($b_i \approx 1$) and deviation from regression non-significantly deviate from zero ($S^2_{di}=0$) with high mean performance over the population mean. The value $b_i \approx 1$ indicates that the genotype is less responsive to the environmental changes and therefore is more adaptive. However, if b_i is less than unity ($b_i < 1$), the genotype is likely to perform well under poor environmental condition and if b_i value more than unity ($b_i > 1$), the genotype will perform better under favourable environment conditions. Significant deviation from regression (S^2_{di}) will invalidate for the linear prediction, whereas non-significant S^2_{di} indicate that the performance of a genotype in a given set of environmental condition may be predicted. Similar findings were reported by Sharma *et al.*, 2012; Arain *et al.*, 2011; Aydin *et al.*, 2010; Seboksa *et al.*, 2001; Akcura *et al.*, 2005 and Arshad *et al.*, 2003.

ANOVA (Table 2) showed that the sources of variation for Environment + (G $\times$ E) was found highly significant. Highly significant ($P < 0.01$) variation was observed in environment and genotype $\times$ environment interaction. A significant ($p < 0.05$) variation is also noted in genotypes. Significance of GEI is an indication for inconsistency of genotypes in response to changing climatic conditions across locations due to genotype-environment interaction. These results are in conformity with Kulkarni *et al.* (2021), Kulkarni *et al.* (2022), Das *et al.* (2010), Tiawari *et al.* (2011) and Zerihun (2011), Matin *et al.* (2017). Higher variation was observed across environment similarly as reported by Letta (2009) and Das *et al.* (2010). The magnitude of variation instigated by environment reveals that complex external factors (biotic and abiotic) are number one challenges in crop improvement. The larger variation observed in GEI, discriminate the correlation between phenotype and genotype, hence, it is hard to

evaluate the genetic potential of particular genotype whose performance changed in different environments.

Considering the overall performance, UHSBC_67 (G4) was found to be promising and well adapted to all environments across seasons for shoot length, root weight, and shoot weight. Therefore, UHSBC_67 can be used for cultivation across seasons for these traits. UHSBC_117 (G3) was stable for root length and root width. Both UHSBC_117 (G3), UHSBC_67 (G4) was found to be unstable for shoulder width. UHSBC_32-2 (G1), UHSBC_117 (G3), UHSBC_23-1 (G5) and UHSBC_100 (G7) exhibited suitability under unfavourable environment for root width, root weight, number of petioles traits. Genotype UHSBC_34-1 (G6) was found to be stable genotype for shoot length, shoot weight and number of petioles and root width under favourable environment but we could observe that same genotype was found to be unstable for root length, root weight and shoulder width. None of the carrot cultivars were stable across changing environments across seasons of northern dry zone of Karnataka evaluated under present study. The change in monsoon and weather environments during crop growth period could be the major reason for instability of carrot performance across seasons. Hence, the present work highlights the necessity of understanding ecological effect on carrot crop growth and emphasizes the need for development of environment specific cultivars to tropical climates.

Evaluation of genotypes across locations revealed that, UHSBC_34-1 (G6) exhibited non-significant deviation from regression and regression coefficient less than unity ($b_i < 1$) with mean value close to the population for all traits except shoot length indicating its stability under unfavourable environments. Whereas, UHSBC_100(G7), showed stable performance across all environments for shoot length, number of petioles, shoot weight, root weight. Genotype UHSBC_23-1(G5), was found stable for number of petioles, shoot weight and root weight across all environments. Thus, these genotypes can be used for general cultivation across all environments. UHSBC_117(G3), for shoot length, shoulder width and root weight and UHSBC_67 (G4), for shoot weight, shoulder width and root weight traits, showing stability under favourable environments showing non-significant deviation from regression and

regression coefficient around unity ($b_i > 1$) along with mean value close to the population mean.

UHSBC_32-2(G1), UHSBC_17(G2), UHSBC_117(G3), UHSBC_67(G4) and UHSBC_23-1(G5) genotypes were having unit regression coefficient ($b_i \approx 1$) and deviation from regression non-significantly deviate from zero ($S^2_{di}=0$) with high mean performance over the population mean for root length across all environments, thereby indicating their stability and adaptability under all environments. Data further revealed that, UHSBC_67 (G4) genotype had stability for shoot length and number of petioles as well across three different environments.

None of the carrot cultivars were stable across changing environments across three locations and three seasons for all seven traits under northern dry zone of Karnataka. Climatic factors, as well as soil conditions play prominent role in the vegetative growth period of carrot. Therefore, the present work reveals the necessity of understanding $G \times E$ interactions and its effect on co-inheritance of traits for development of environment specific cultivars to tropical climates. $G \times E$ interaction, coupled with high mean performance indicate that the genotypes are suitable for general adaptation in the range of environments considered.

Environmental index and phenotypic index helped to understand the impact of environments such as seasons and locations on the performance of carrots genotypes. Specific location (Bagalkote) and season (Year 2017 and year 2019) found to be favourable environments. It was also shown that temperate adapted carrot cultivar used to evaluate across seasons and locations was least responsive to varied environmental conditions compared to tropical adapted carrot cultivars as shown by negative PI. In this regard, efforts are also necessary to utilize the adaptive mechanism of tropical carrot cultivars in temperate types by hybridization method for increasing the temperate or tropical carrot crop growing area in the context of climate resilient. This would further provide the privilege to consumers and farmers of extreme climatic conditions for consumption of this more nutritious vegetable crop in terms of raw or processed form.

With the present understanding of transcriptome studies on these two contrasting and economically important orange carrot types, it was possible to understand the molecular mechanism underlying root plasticity, root adaptability, innate immune response to various biotic and abiotic factors, contrasting flowering requirements *etc* in a broad perspective. The identified DEGs, transcriptional factors, genetic variants of the present study are beneficial to carrot breeders, molecular biologists for exploring them in carrot crop improvement. The vast genomic resources generated in the present study are of great utility in the trait targeted breeding or overall crop improvement program for long time in future. There is a need to validate few more specific genes and transcriptional factors along with the genetic variants identified in the present study for their applicability in precise breeding like genomic assisted breeding and/ genome editing.

6. SUMMARY AND CONCLUSION

Carrot is a biennial crop; its life cycle consists of two phases viz., vegetative and reproductive/generative phases. Seed is the propagating material and secondary storage taproot root is the edible part, both the phases are considered to be important for farmers and breeders. However, a greater effort has been made on understanding the nutritional value, genetics and genomics of only temperate adapted, 'Western orange' carrot cultivar despite the availability of germplasm such as 'Asiatic orange cultivars' that adapt to wide range of environments.

Carrot production is generally restricted to cool climate for seed production leading to non-affordable seed price to tropical and subtropical farmers, sequentially for consumers to buy edible root as a vegetable with the affordable rate. Selection pressure has improved beneficial traits like, greater provitamin A carotenoid and higher sugar content in a highly domesticated Western carrots than Asiatic carrot. Greater diversity was found in our study for these two cultivated carrot types in both root development and flowering biology. Western types are rich sources of β -carotenoid (>10.0 mg/100gm), have higher root uniformity and fetch higher market price for the farmers. However, Asiatic types are better adaptable to wide climates, withstand biotic and abiotic stresses and more importantly can able to produce flowering, despite higher temperature in tropics, making it feasible to produce the seeds as it is the most important and profitable produce for the farmers. However, genetic mechanisms that contribute to phenotypic changes associated with these two contrasting carrot cultivars remains unknown despite the availability of high-quality carrot genome (473Mb).

Hence, the present investigation was designed to elucidate the biology, physiology, genetic and molecular basis of secondary storage root development, adaptation and vegetative to reproductive transition respectively from two economic tissues viz., storage root and floral primordia in two important cultivated carrots i.e. Western and Asiatic types at two phases (vegetative and reproductive) of carrot life cycle by dynamic transcriptome profiling with next generation sequencing (NGS) technology platforms. Flowering was successfully induced in both the cultivars after vegetative root production. As per our knowledge, these high throughput genetic and

molecular comparisons (cultivar types, tissues and developmental phases) are the first comprehensive information in carrot globally.

Based on the phenotypic results it was observed that tropical cultivar showed early flowering behavior with less or no vernalization requirement than temperate cultivar during reproductive stage. Temperate carrot was superior in terms of attractive dark orange color, uniform inner core with no clear xylem or phloem patterning and higher beta carotene in contrast to tropical cultivar, but did not switch from vegetative to reproductive stage under tropics.

Triplicate samples from each of the cultivar and tissues were used to generate twelve RNASeq libraries with the raw reads of >140 million reads for roots and >233 million reads for floral tissues. For computational analysis, reference based (high quality carrot reference genome) as well as *de novo* assembly was developed from TopHat and Trinity tools respectively. High quality assembly was generated in both reference-based with >70% assembly rate and *de novo* assembly having >89% BUSCO completeness and >98% assembly of the individual libraries. Further, identification of differentially expressed genes (DEGs) was performed by cufflinks, DESeq and edgeR tools. A total of 3544 and 2239 DEGs in reference based and *de novo* assemblies respectively were identified in root tissues with 1818 and 1090 up regulated genes in Asiatic and 1726 and 1149 genes up regulated in Western cultivars from reference based and *de novo* assembly with p value <0.05 and log₂ fold change >2. ~44k Unigenes were identified in the present study with 477 total transcriptional factors (TF) with higher number of stress responsive TF were up regulated in Asiatic cultivar (bHLH, ERF, NAC and WRKY) than Western type. In the root transcriptome >45% of the genes were annotated with various annotation tools David, Blastx, Interpro *etc* considering Eudicot lineage.

Gene ontology (GO) categorization of the DEGs into biological process (BP), cellular components (CC), molecular functions (MF) as well as KEGG pathway enriched pathway analysis identified major important functions and terms in each of the category specific to Western and Asiatic cultivars. Major pathways enriched in Western cultivar roots were specific to photosynthesis, glycolysis/gluconeogenesis, 'plant hormone signal transduction', 'carbon metabolism', 'photosynthesis' *etc* that helps in

better source-sink translocation ability and accumulation of higher carotenoid in roots. Whereas, in Asiatic cultivar, enriched pathways included were mainly ‘protein phosphorylation’, plant-pathogen interaction, MAPK signaling pathways, phenyl propanoid biosynthesis indicating better root plasticity adaptation and tolerance to biotic or abiotic stresses.

It was interesting to note that, despite having higher stringency in DEG identification with $\log_2FC$ of ± 2 , some of these terms were commonly identified in reference and *de novo*-based DEG identification. The results clearly indicate the significance and functionality of gene or gene products in carrot growth, development or genetic makeup as well as the reliability of the RNA-seq data in distinguishing and exhibiting the importance of these two contrasting cultivated carrot types.

A total of 24 pathways were mapped to the KEGG database in reference based DEGs where in 15 were upregulated in tropical cultivar/downregulated in temperate and 9 were upregulated in temperate cultivar or downregulated in tropical. Whereas, in tropical upregulated pathways, other important KEGG mapped pathways was plant-pathogen interaction, MAPK signalling pathways, phenyl propanoid biosynthesis, *etc.*

Gene regulatory network (GRN) analysis of root transcriptome also confirmed the co-expression of photosynthesis and vascular cambium related genes in association with stress responsive genes in Western cultivar and Asiatic cultivars and are considered as key factors that distinguish these two contrasting cultivar types of carrot for storage root development. Understanding the genetic architecture of this plasticity may possibly provide useful breeding targets for crop improvement specifically in tropical or sub-tropical environments that constitute the major part of the cultivation area of globe. GRN for flowering-related genes for FPL tissue clearly demonstrated the co-expression of key flowering-related genes in tropical and temperate cultivars in different levels of expression.

The putative differentially expressed candidate genes responsible for root adaptability, root (GO:0010051, GO0010588 *etc.*), carbon and lipid metabolism (GO:0051762, GO:0016117, GO:0016126), photosynthetic cascades (LSH10, PSAH, PSAN, PSAD, LHCB, PSB, *etc.*), biotic and abiotic stress-related genes (WRKY70, ERF-71, MKK9, MYB15, *etc.*), vascular cambium patterning (LSH10, PSAH, PSAN,

PSAD, LHCB, PSB), and flower development (GI, CO, PhyA, AP2, FKF1, FT) of the present study would be of great utility in development of ‘superior tropical varieties’ by ‘genomic assisted breeding’ in carrot and other crops of similar relevance. The technology to induce flowering Western cultivar could be a breakthrough technology useful for tropical farmers and seed industries to increase the carrot growing area.

Validation of DEGs of root and FPL tissues of RNA-seq data using qPCR results confirms the quality and reliability of the generated data for their direct utilization in carrot breeding. Further, genes validated for important flowering genes in temperate and tropical cultivars would be utilized to elucidate the mysterious flowering pathway operating in these two domesticated orange cultivars contrasting for flowering habit for proper exploration of both the gene pools to increase the global carrot production in the years to come.

From the stability analysis, environmental index and phenotypic index helped to understand the impact of environments such as seasons and locations on the performance of carrots genotypes. Specific location (Bagalkote) and season (Year 2017 and year 2019) found to be favorable environments. It was also shown that temperate adapted carrot cultivar used to evaluate across seasons and locations was least responsive to varied environmental conditions compared to tropical adapted carrot cultivars as shown by negative PI. In this regard, efforts are also necessary to utilize the adaptive mechanism of tropical carrot cultivars in temperate types by hybridization method for increasing the temperate or tropical carrot crop growing area in the context of climate resilient. This would further provide the privilege to consumers and farmers of extreme climatic conditions for consumption of this more nutritious vegetable crop in terms of raw or processed form.

Outcomes of the present investigation

- First report on comparative transcriptome profiling of two economically important contrasting adaptive carrot cultivars belonging to orange types with different genes pool for two economic tissues at two phases of carrot life cycle for molecular understanding of genes and TF.

- The flowering induction strategy could be useful for breeders to undertake hybridization program for improvement of domesticated traits like root colour, nutrient content from temperate to tropical and stress responsive genes from tropical to temperate cultivars.
- It is also beneficial to seed industries to increase global carrot production and for farmers to produce their own seeds after root production so that they can get double profit from root and seeds.
- As per the current literature, it is a first RNA-seq experiment to understand the expression profiling by ‘comparative transcriptomics’ of Asiatic/Temperate and Western/tropical adapted carrots in two economic tissues at two developmental phases of carrot (root+ floral primordia).
- Gene Regulatory Network understanding helps reduce the undesirable core in Asiatic carrots by genetically engineering the vascular cambium related master regulatory genes
- Huge RNA-seq data generate in the present study is of great utility for studying domestication related traits, root adaptability- biotic and abiotic stress tolerance and their transfer to the desirable cultivars
- Flowering pathway key genes are useful for immediate utility in crop breeding
- Candidate genes for root plasticity, biotic and abiotic stress tolerance, flowering traits and nutritionally important genes can be directly used for “Genomic assisted selection”
- Genetic variants (SSRs- *indels* and SNPs in root and FPL) identified in the present study are useful in QTL mapping, diversity assessment and others

Future prospects of present research

- It is necessary to optimize the flowering in temperate adapted Western cultivar by adopting a technology considering all the flowering pathways based on the current research to produce seeds under tropical region.

- Development of ‘diagnostic markers’ for candidate genes identified in the present study for genomic selection or ‘genomic assisted breeding’
- Based on the current understanding of vascular cambium patterning, attempt must be made to make uniformity of the phloem pigmentation by reducing the xylem core thickness either by traditional selection or advanced genetic engineering methods.
- There is a need to identify polymorphic genetic variants and functional annotation for effective utilization of the identified genetic variants for mapping and MAS.

REFERENCES

- Ahmad, T., Cawood, M., Iqbal, Q., Ariño, A., Batool, A., Tariq, R.M.S., Azam, M. and Akhtar, S., 2019, Phytochemicals in *Daucus carota* and their health benefits. *Foods*, 8(9): 424.
- Ajani, Y., Bull-Herenu, K. and Claben-Bockhoff, R., 2016, Patterns of flower development in Apiaceae–Apioidae. *Flora*, 221: 38-45.
- Akcura, M., Ceri, S., Taner, S., Kaya, Y., Ozer, E. and Ayranci, R., 2005, Grain yield stability of winter oat (*Avena sativa* L.) cultivars in the central Anatolian region of Turkey. *J. Cent. Eur. Agric.*, 6(3): 203-210.
- Albert, E.V., Kawai-ool, U.N. and Ezhova, T.A., 2014, Gene NANA regulates cell proliferation in *Arabidopsis thaliana* shoot apical meristem without interaction with CLV1, CLV2, CLV3. *Russ. J. Dev. Biol.*, 45(5): 267-272.
- Alessandro, M.S. and Galmarini, C.R., 2007, Inheritance of vernalization requirement in carrot. *J. Am. Soc. Hortic. Sci.*, 132(4): 525-529.
- Alessandro, M.S., Galmarini, C.R., Iorizzo, M. and Simon: W., 2013, Molecular mapping of vernalization requirement and fertility restoration genes in carrot. *Theor. Appl. Genet.*, 126(2): 415-423.
- Alexeyenko, A. and Sonnhammer, E.L., 2009, Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res.*, 19(6): 1107-1116.
- Ali, A., Matthews, W.C., Cavagnaro: F., Iorizzo, M., Roberts: A. and Simon: W., 2014, Inheritance and mapping of Mj-2, a new source of root-knot nematode (*Meloidogyne javanica*) resistance in carrot. *J. Hered.*, 105(2): 288-291.
- Alibés, A., Yankilevich: and Díaz-Uriarte, R., 2007, ID converter and IDCligh: conversion and annotation of gene and protein IDs. *BMC bioinformatics*, 8(1): 1-9.

- Amasino, R., 2004, Vernalization, competence, and the epigenetic memory of winter. *Plant Cell*, 16(10): 2553-2559.
- Amasino, R., 2010, Seasonal and developmental timing of flowering. *Plant J.*, 61(6): 1001-1013.
- Amasino, R.M., Michaels, S.D., 2010, The timing of flowering. *Plant Physiol.*, 154: 516–520.
- An, H., Roussot, C., Suárez-López, Corbesier, L., Vincent, C., Piñeiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C. and Coupland, G., 2004, Constans acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*.
- Anders, S., Pyl: T. and Huber, W., 2015, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2): 166-169.
- Andrés, F. and Coupland, G., 2012, The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.*, 13(9): 627-639.
- Andrews, S., 2010, FastQC: a quality control tool for high throughput sequence data.
- Angell, F. F., and Gabelman, W. H., 1968, Inheritance of resistance in carrot *Daucus carota* var *sativa* to leaf spot fungus *Cercospora carotae*. In *proceedings Amer. Soc. Horti. Sci.*, 93: 434-701.
- Angell, F.F. and Gabelman, W.H., 1970, Inheritance of purple petiole in carrot, *Daucus carota* var. *sativa*. *Hortic. Sci.*, 5:175.
- Arbizu, C., Ruess, H., Senalik, D., Simon: W. and Spooner, D.M., 2014, Phylogenomics of the carrot genus (*Daucus*, *Apiaceae*). *Am. J. Bot.*, 101(10): 1666-1685.
- Arbizu, C.I., Ellison, S.L., Senalik, D., Simon: W. and Spooner, D.M., 2016a, Genotyping-by-sequencing provides the discriminating power to investigate the subspecies of *Daucus carota* (*Apiaceae*). *BMC Evol. Biol.*, 16(1):1-16.

- Arbizu, C.I., Simon: W., Martínez-Flores, F., Ruess, H., Crespo, M.B. and Spooner, D.M., 2016b. Integrated molecular and morphological studies of the *Daucus guttatus* complex (*Apiaceae*). *Systematic Botany*, 41(2): 479-492.
- Arain, M.A., Sial, M.A., Rajput, M.A. and Mirbahar, A.A., 2011. Yield stability in bread wheat genotypes. *Pak. J. Bot*, 43(4), pp.2071-2074.
- Arias, D.B., Pinto, K.A.G., Cooper, K.K. and Summers, M.L., 2020. Transcriptomic analysis of cyanobacterial alkane overproduction reveals stress-related genes and inhibitors of lipid droplet formation. *Microbial genomics*, 6(10).
- Arshad, M., Bakhsh, A., Haqqani, A.M. and Bashir, Muhammad, 2003, Genotype-environment interaction for grain yield in chickpea (*Cicer arietinum* L.). *Pak. J. Bot*, 35(2):181-186.
- Arumuganathan, K. and Earle, E.D., 1991, Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.*, 9(3): 208-218.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. and Harris, M.A., 2000, Gene ontology: tool for the unification of biology. *Nat. Genet.*, 25(1): 25-29.
- Atherton, J.G., Basher, E.A. and Brewster, J.L., 1984, The effects of photoperiod on flowering in carrot. *J. Hortic. Sci.*, 59(2): 213-215.
- Atherton, J.G., Craigon, J. and Basher, E.A., 1990. Flowering and bolting in carrot. I. Juvenility, cardinal temperatures and thermal times for vernalization. *J. Hortic. Sci.*, 65(4): 423-429.
- Atkins, M., Potier, D., Romanelli, L., Jacobs, J., Mach, J., Hamaratoglu, F., Aerts, S. and Halder, G., 2016, An ectopic network of transcription factors regulated by hippo signaling drives growth and invasion of a malignant tumor model. *Curr. Biol.*, 26(16): 2101-2113.
- Ausin, I., Alonso-Blanco, C. and Martinez-Zapater, J.M., 2004, Environmental regulation of flowering. *Int. J. Dev. Biol.*, 49(5-6): 689-705.

- Aydin, N., Bayramoglu, H. O. and Ozcan, F.L. (2010). Stability of some quality traits in bread wheat (*Triticum aestivum*) genotypes. *Journal of Environment Biology*, 31: 489-95.
- Baggett, J.R. and Kean, D., 1989, Inheritance of annual flowering in *Brassica oleracea*. *Hort. Sci.*, 24(4): 662-664.
- Banakar, Hada, A., Papolu: K. and Rao, U., 2020, Simultaneous RNAi knockdown of three FMRFamide-like peptide genes, Mi-flp1, Mi-flp12, and Mi-flp18 provides resistance to root-knot nematode, *Meloidogyne incognita*. *Front. Microbiol.*, 11: 2690.
- Banasiak, Ł., Wojewódzka, A., Baczyński, J., Reduron, J.P., Piwczyński, M., Kurzyna-Młynik, R., Gutaker, R., Czarnocka-Cieciura, A., Kosmala-Grzechnik, S. and Spalik K., 2016, Phylogeny of Apiaceae subtribe Daucinae and the taxonomic delineation of its genera. *Taxon*. 65: 563–585
- Banga, O., 1957, The development of the original European carrot material. *Euphytica*. 6(1): 64-76.
- Banga, O., Petiet, J. and Van Bennekom, J.L., 1964, Genetical analysis of male-sterility in carrots, *Daucus carota* L. *Euphytica*, 13(1): 75-93.
- Bannenberg, G., Martínez, M., Hamberg, M. and Castresana, C., 2009, Diversity of the enzymatic activity in the lipoxygenase gene family of *Arabidopsis thaliana*. *Lipids*, 44(2): 85.
- Bao, F., Shen, J., Brady, S.R., Muday, G.K., Asami, T. and Yang, Z., 2004, Brassinosteroids interact with auxin to promote lateral root development in *Arabidopsis*. *Plant Physiol.*, 134(4): 1624-1631.
- Baranski, R., Maksylewicz-Kaul, A., Nothnagel, T., Cavagnaro: F., Simon: W. and Grzebelus, D., 2012, Genetic diversity of carrot (*Daucus carota* L.) cultivars revealed by analysis of SSR loci. *Genet. Resour. Crop Evol.*, 59(2): 163-170.

- Bäurle, I. and Dean, C., 2006, The timing of developmental transitions in plants. *Cell*, 125(4): 655-664.
- Becerra-Moreno, A., Redondo-Gil, M., Benavides, J., Nair, V., Cisneros-Zevallos, L. and Jacobo-Velázquez, D.A., 2015, Combined effect of water loss and wounding stress on gene activation of metabolic pathways associated with phenolic biosynthesis in carrot. *Front. Plant Sci.*, 6: 837-841.
- Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M., 2017, MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33(16): 2583-2585.
- Bell, E., Creelman, R.A. and Mullet, J.E., 1995, A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in *Arabidopsis*. *Proc. Natl. Acad. Sci.*, 92(19): 8675-8679.
- Benjamini, Y. and Hochberg, Y., 1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Society: series B (Methodological)*, 57(1): 289-300.
- Bianchi, V.J., Rubio, M., Trainotti, L., Verde, I., Bonghi, C. and Martínez-Gómez, , 2015, Prunus transcription factors: breeding perspectives. *Front. Plant Sci.*, 6: 443.
- Bianchi, V.J., Rubio, M., Trainotti, L., Verde, I., Bonghi, C. and Martínez-Gómez, , 2015, Prunus transcription factors: breeding perspectives. *Front. Plant Sci.*, 6: 443.
- Bolger, A.M., Lohse, M. and Usadel, B., 2014, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120.
- Bonnet, A., 1983, Source of resistance to powdery mildew for breeding cultivated carrots. *Agronomie* 3:33–37
- Börner, T., Linke, B., Nothnagel, T., Scheike, R., Schulz, B., Steinborn, R., Brennicke, A., Stein, M. and Wricke, G., 1995, Inheritance of nuclear and cytoplasmic factors affecting male sterility in *Daucus carota*. *Fortschritte der Pflanzenzuechtung* (Germany).

- Bostan, H. and Chiusano, M.L., 2015, NexGenEx-Tom: a gene expression platform to investigate the functionalities of the tomato genome. *BMC Plant Boil.*, 15(1): 1-13.
- Bostan, H., Senalik, D., Simon: W. and Iorizzo, M., 2019, Carrot genetics, omics and breeding toolboxes. In *The Carrot Genome* (pp. 225-245). Springer, Cham.
- Braak, J.P. and Kho, Y.O., 1958, Some observations on the floral biology of the carrot (*Daucus carota* L.). *Euphytica*, 7(2): 131-139.
- Bradeen, J.M. and Simon: W., 1998, Conversion of an AFLP fragment linked to the carrot Y2 locus to a simple, codominant, PCR-based marker form. *Theor. Appl. Genet.*, 97(5-6): 960-967.
- Bradeen, J.M., Bach, I.C., Briard, M., Le Clerc, V., Grzebelus, D., Senalik, D.A. and Simon: W., 2002. Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. *J. Am. Soc. Hortic. Sci.*, 127(3): 383-391.
- Bremer, B., 2009, In: Hedges SB, Kumar S (eds) The time tree of life. Oxford University Press, New York, pp 177–187.
- Brian W., 1959, Effects of gibberellins on plant growth and development. *Biol. Rev.*, 34(1): 37-77.
- Broussard, M.A., Mas, F., Howlett, B., Pattemore, D. and Tylianakis, J.M., 2017, Possible mechanisms of pollination failure in hybrid carrot seed and implications for industry in a changing climate. *PloS one*, 12(6): e0180215.
- Budahn, H., Barański, R., Grzebelus, D., Kielkowska, A., Straka: , Metge, K., Linke, B. and Nothnagel, T., 2014, Mapping genes governing flower architecture and pollen development in a double mutant population of carrot. *Front. Plant Sci.*, 5: 504.
- Buishand, J.G. and Gabelman, W.H., 1979. Investigations on the inheritance of color and carotenoid content in phloem and xylem of carrot roots (*Daucus carota* L.). *Euphytica*, 28(3): 611-632.

- Bustin, S.A., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G. and Wittwer, C.T., 2013. The need for transparency and good practices in the qPCR literature. *Nature methods*, **10**(11): 1063-1067.
- Caldelari, D., Wang G., Farmer, E.E. and Dong X., 2011, *Arabidopsis* lox3 lox4 double mutants are male sterile and defective in global proliferative arrest. *Plant Mol. Biol.*, 75(1): 25-33.
- Cao, F.Y., Yoshioka, K. and Desveaux, D., 2011, The roles of ABA in plant–pathogen interactions. *J. Plant Res.*, 124(4): 489-499.
- Capovilla, G., Schmid, M. and Posé, D., 2015, Control of flowering by ambient temperature. *J. Exp. Bot.*, 66(1): 59-69.
- Carianopol, C.S., Chan, A.L., Dong S., Provart, N.J., Lumba, S. and Gazzarrini, S., 2020, An abscisic acid-responsive protein interaction network for sucrose non-fermenting related kinase1 in abiotic stress response. *Commun. Biol.*, 3(1): 1-15.
- Cavagnaro: F., 2019. Genetics and genomics of carrot sugars and polyacetylenes. In *The Carrot Genome Springer*, Cham. (pp. 295-315).
- Cavagnaro: F., Chung S.M., Manin, S., Yildiz, M., Ali, A., Alessandro, M.S., Iorizzo, M., Senalik, D.A. and Simon: W., 2011, Microsatellite isolation and marker development in carrot-genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across Apiaceae. *BMC genomics*, 12(1): 1-20.
- Cavagnaro: F., Chung S.M., Szklarczyk, M., Grzebelus, D., Senalik, D., Atkins, A.E. and Simon: W., 2009, Characterization of a deep-coverage carrot (*Daucus carota* L.) BAC library and initial analysis of BAC-end sequences. *Mol. Genet. Genom.*, 281(3): 273-288.

- Cavagnaro: F., Iorizzo, M., Yildiz, M., Senalik, D., Parsons, J., Ellison, S. and Simon: W., 2014, A gene-derived SNP-based high resolution linkage map of carrot including the location of QTL conditioning root and leaf anthocyanin pigmentation. *BMC genomics*, 15(1): 1-17.
- Celaj, A., Markle, J., Danska, J. and Parkinson, J., 2014, Comparison of assembly algorithms for improving rate of meta transcriptomic functional annotation. *Microbiome*, 2(1): 1-13.
- Chaitra, 2017, Genetic Diversity, Population Structure and Marker-Trait Association for Root Traits by Association Analysis in Carrot (*Daucus carota* L.). M.Sc Thesis submitted to UHS, Bagalkot
- Chaitra, K.C., Sarvamangala, C., Manikanta, D.S., Chaitra: A. and Fakrudin, B., 2020, Insights into genetic diversity and population structure of Indian carrot (*Daucus carota* L.) accessions. *J. Appl. Genet.*, 61(3): 303-312.
- Chakraborty, S., Nguyen, B., Wasti, S.D. and Xu, G., 2019, Plant leucine-rich repeat receptor kinase (LRR-RK): structure, ligand perception, and activation mechanism. *Mol.*, 24(17): 3081.
- Challa, G., 2018, Transcriptome analysis of root development in wheat *Triticum aestivum* using high throughput sequencing technologies. *Electronic Theses and Dissertations*. 2655.
- Chauvin, A., Caldelari, D., Wolfender, J.L. and Farmer, E.E., 2013, Four 13 - lipoxygenases contribute to rapid jasmonate synthesis in wounded *Arabidopsis thaliana* leaves: a role for lipoxygenase 6 in responses to long - distance wound signals. *New Phytol.*, 197(2): 566-575.
- Chen, D., Yan, W., Fu, L.Y. and Kaufmann, K., 2018, Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. *Nat. Commun.*, 9(1): 1-13.

- Chen, Y., Souaiaia, T. and Chen, T., 2009, PerM: efficient mapping of short sequencing reads with periodic full sensitive spaced seeds. *Bioinformatics*, 25(19): 2514-2521.
- Chialva, C., Blein, T., Crespi, M. and Lijavetzky, D., 2021, Insights into long non-coding RNA regulation of anthocyanin carrot root pigmentation. *Sci. Rep.*, 11(1): 1-10.
- Choi, K., Kim, J., Hwang, H.J., Kim, S., Park, C., Kim, S.Y. and Lee, I., 2011, The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. *Plant Cell*, 23(1): 289-303.
- Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C.E. and Burow, M.D., 2014, Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. *PloS one*, 9(12): e115055.
- Claverie, M., Dirlwanger, E., Bosselut, N., Van Ghelder, C., Voisin, R., Kleinhentz, M., Lafargue, B., Abad, , Rosso, M.N., Chalhoub, B. and Esmenjaud, D., 2011, The Ma gene for complete-spectrum resistance to *Meloidogyne* species in *Prunus* is a TNL with a huge repeated C-terminal post-LRR region. *Plant Physiol.*, 156(2): 779-792.
- Clotault, J., Geoffriau, E., Lionneton, E., Briard, M. and Peltier, D., 2010, Carotenoid biosynthesis genes provide evidence of geographical subdivision and extensive linkage disequilibrium in the carrot. *Theor. Appl. Genet.*, 121(4): 659-672.
- Cock, J., Fields, C.J., Goto, N., Heuer, M.L. and Rice, M., 2010, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.*, 38(6): 1767-1771.
- Coenye, T., 2021, Do results obtained with RNA-sequencing require independent verification. *Biofilm*, 3.

- Conesa, A., Gotz, S., García-Gómez, J.M., Terol, J., Talon, M. and Robles, M., 2005, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18): 3674-3676.
- Consortium, U., 2015, UniProt: a hub for protein information. *Nucleic Acids Res.*, 43(1): 204-212.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G., 2007, FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science*, 316(5827): 1030-1033.
- de Jong, T.J., Grebenstein, C. and Tamis, W.L., 2016, Demography and life history of *Daucus carota* L. populations in the Netherlands. *Flora*, 224: 154-158.
- Den Herder, G., Van Isterdael, G., Beeckman, T. and De Smet, I., 2010, The roots of a new green revolution. *Trends Plant Sci.*, 15(11): 600-607.
- Dennis, E.S., Finnegan, E.J., Bilodeau, , Chaudhury, A., Genger, R., Helliwell, C.A., Sheldon, C.C., Bagnall, D.J. and Peacock, W.J., 1996, June. Vernalization and the initiation of flowering. In *Seminars in Cell and Developmental Biology* (Vol. 7, No. 3, pp. 441-448). Academic Press.
- Das, S., Misra, R.C., Patnaik, M.C. and Das, S.R. GxE interaction, adaptability and yield stability of mid-early rice genotypes. *Indian journal of agricultural research*, 2010; 44(2):104-111.
- Dias Tagliacozzo, G.M. and Valio, I.F.M., 1994, Effect of vernalization on flowering of *Daucus carota* (Cvs Nantes and Brasilia). *Rev. Bras. Fisiol. Veg.*, 6(1): 71-73.
- Díaz-Martín, J., Almoguera, C., Prieto-Dapena, , Espinosa, J.M. and Jordano, J., 2005, Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. *Plant Physiol.*, 139(3): 1483-1494.

- Dickson, M.H. and Peterson, C.E., 1958. Hastening greenhouse seed production for carrot breeding. *In Proc. Amer. Soc. Hort. Sci.*, (Vol. 71, No. 5, pp. 412-415).
- Dickson, M.H., 1966, The inheritance of longitudinal cracking in carrots. *Euphytica*, 15(1): 99-101.
- Ding, Z., Li, S., An, X., Liu, X., Qin, H. and Wang, D., 2009, Transgenic expression of MYB15 confers enhanced sensitivity to abscisic acid and improved drought tolerance in *Arabidopsis thaliana*. *J. Genet. Genomics*, 36(1): 17-29.
- Doebley, J.F., Gaut, B.S. and Smith, B.D., 2006, The molecular genetics of crop domestication. *Cell*. 127: 1309–1321
- Dong, A.X., Xin, H.B., Li, Z.J., Liu, H., Sun, Y.Q., Nie, S., Zhao, Z.N., Cui, R.F., Zhang, R.G., Yun, Q.Z. and Wang, X.N., 2018, High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *Giga Science*, 7(7): 68.
- Drisch, R.C. and Stahl, Y., 2015, Function and regulation of transcription factors involved in root apical meristem and stem cell maintenance. *Front. Plant Sci.*, 6: 505.
- Du, D., Hao, R., Cheng, T., Pan, H., Yang, W., Wang, J. and Zhang, Q., 2013, Genome-wide analysis of the AP2/ERF gene family in *Prunus mume*. *Plant Mol. Biol. Rep.*, 31(3): 741-750.
- Du, J., Li, M., Yuan, Z., Guo, M., Song, J., Xie, X. and Chen, Y., 2016, A decision analysis model for KEGG pathway analysis. *BMC bioinformatics*, 17(1): 1-12.
- Eberhart, S.T. and Russell, W.A., 1966, Stability parameters for comparing varieties 1. *Crop Sci.*, 6(1): 36-40.
- Ekblom, R. and Wolf, J.B., 2014, A field guide to whole - genome sequencing, assembly and annotation. *Evol. Appl.*, 7(9): 1026-1042.

- Ellison, S., 2019, Carrot Domestication. In *The Carrot Genome*; Simon: W., Iorizzo, M., Grzebelus, D., Baranski, R., Eds.; Springer: Cham, Switzerland, pp. 77–91.
- Ellison, S., Bostan, H., Simon, Senalik, D. and Iorizzo, M., 2017, Fine mapping transcriptome analysis, and marker development for, the gene that conditions β^2 -carotene accumulation in Carrot (L.). *Genes Genomes Genetics*.
- Ellison, S.L., Luby, C.H., Corak, K.E., Coe, K.M., Senalik, D., Iorizzo, M., Goldman, I.L., Simon: W. and Dawson, J.C., 2018. Carotenoid presence is associated with the *or* gene in domesticated carrot. *Genetics*, 210(4): 1497-1508.
- Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S. and Weinstock, G.M., 2007, Creating a honey bee consensus gene set. *Genome Biol.*, 8(1): 1-8.
- Erbar, C., Leins, van Wyk, B.E. and Tilney: M., 2004, Sympetaly in Apiales (Apiaceae, Araliaceae, Pittosporaceae). *Sou. Afr. J. Bot.*, 70(3): 458-467.
- Eulgem, T., Rushton: J., Robatzek, S. and Somssich, I.E., 2000, The WRKY superfamily of plant transcription factors. *Trends Plant Sci.*, 5(5): 199-206.
- Eulgem, T., Rushton: J., Robatzek, S. and Somssich, I.E., 2000, The WRKY superfamily of plant transcription factors. *Trends Plant Sci.*, 5(5): 199-206.
- Ewels, Magnusson, M., Lundin, S. and Käller, M., 2016, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19): 3047-3048.
- FAO, 2017, The state of food security and nutrition in Africa: building resilience for peace and food security. Rome.
- Ferreira, M.E., Satagopan, J., Yandell, B.S., Williams: H. and Osborn, T.C., 1995, Mapping loci controlling vernalization requirement and flowering time in *Brassica napus*. *Theor. Appl. Genet.*, 90(5): 727-732.

- Finn, R.D., Coghill, Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. and Salazar, G.A., 2016, The Pfam protein family's database: towards a more sustainable future. *Nucleic Acids Res.*, 44(1): 279-285.
- Firon, N., LaBonte, D., Villordon, A., Kfir, Y., Solis, J., Lapis, E., Perlman, T.S., Doron-Faigenboim, A., Hetzroni, A., Althan, L. and Nadir, L.A., 2013, Transcriptional profiling of sweet potato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC genomics*, 14(1): 1-25.
- Fisher, J.E., 1956, Studies on the photoperiodic and thermal control of flowering in carrots. *Plant Physiol*, 31: 36.
- Flowers, T.J., 2004, Improving crop salt tolerance. *J. Exp. Bot.*, 55(396): 307-319.
- Fornara, F., de Montaigu, A. and Coupland, G., 2010, Snap Shot: control of flowering in *Arabidopsis*. *Cell*, 141(3): 550-550.
- Freeman, R.E., Simon: W., 1983, Evidence for simple genetic control of sugar type in carrot (*Daucus carota* L.). *J. Am. Soc. Hortic. Sci.*, 108: 50–54.
- Galmarini, C.R., Borgo, R. and Tizio, R., 1992, Determination of a pre-vernalization phase in carrot (*Daucus carota* L.) cv. Flakkee. *Turrialba*, 42: 140–142.
- Galmarini, C.R., Della, Gaspara: , 1996, Determinación de requerimientos de pre-vernalización en zanahorias (*Daucus carota* L.) anuales. In: *Actas de la XXI Reunión Argentina de Fisiología Vegetal, Mendoza, Argentina*, p 82
- Gangappa, S.N. and Botto, J.F., 2014, The BBX family of plant transcription factors. *Trends Plant Sci.*, 19(7): 460-470.
- Garner, W.W. and Allard, H.A., 1930, Photoperiodic response of soybeans in relation to temperature and other environmental factors. United States Department of Agriculture.

- Ge, S.X., Jung, D. and Yao, R., 2020, ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8): 2628-2629.
- Ge, S.X., Son, E.W. and Yao, R., 2018, iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* 19, 534. <https://doi.org/10.1186/s12859-018-2486-6>
- Gene Ontology Consortium, 2006, The gene ontology (GO) project in 2006, *Nucleic Acids Res.*, 34: 322-326.
- Glass, K. and Girvan, M., 2014, Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets. *Sci. Rep.*, 4(1): 1-9.
- Glauser, G., Dubugnon, L., Mousavi, S.A., Rudaz, S., Wolfender, J. L. and Farmer, E. E., 2009, Velocity estimates for signal propagation leading to systemic jasmonic acid accumulation in wounded *Arabidopsis*. *J. Biol. Chem.*, 284(50): 34506-34513.
- Glazinska, Wojciechowski, W., Kulasek, M., Glinkowski, W., Marciniak, K., Klajn, N., Kesy, J. and Kopcewicz, J., 2017, De novo transcriptome profiling of flowers, flower pedicels and pods of *Lupinus luteus* (yellow lupine) reveals complex expression changes during organ abscission. *Front. Plant Sci.*, 8:641-647.
- Gohar, T., Payal, A., Grant, M. and Anil, K., 2010, MAPK machinery in plants: recognition and response to different stresses through multiple signal transduction pathways. *Plant Signal. Behav.*, 5(11): 1372-1380.
- Goldman, I. L. and Breitbach, D.N., 1996, Inheritance of a recessive character controlling reduced carotenoid pigmentation in carrot (*Daucus carota* L.). *J. Hered.*, 87(5): 380-382.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., 2011, Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, 29(7): 644.

- Grebner, W., Stingl, N.E., Oenel, A., Mueller, M.J. and Berger, S., 2013, Lipoxygenase6-dependent oxylipin synthesis in roots is required for abiotic and biotic stress resistance of *Arabidopsis*. *Plant Physiol.*, 161(4): 2159-2170.
- Grzebelus, D., 2019, Genetics and genomics of carrot abiotic stress. In *The Carrot Genome* (pp. 363-372). Springer, Cham.
- Grzebelus, D., Baranski, R., Spalik, K., Allender, C. and Simon: W., 2011, *Daucus*. In *Wild crop relatives: genomic and breeding resources* (pp. 91-113). Springer, Berlin, Heidelberg.
- Grzebelus, D., Iorizzo, M., Senalik, D., Ellison, S., Cavagnaro, Macko-Podgorni, A., Heller-Uszynska, K., Kilian, A., Nothnagel, T., Allender, C. and Simon: W., 2014, Diversity, genetic mapping, and signatures of domestication in the carrot (*Daucus carota* L.) genome, as revealed by Diversity Arrays Technology (DArT) markers. *Mol. Breed.*, 33(3): 625-637.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood: D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. and MacManes, M.D., 2013, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, 8(8): 1494-1512.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood: D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. and MacManes, M.D., 2013, De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protoc.*, 8(8): 1494-1512.
- Hammer, K., 1984. Das domestikationssyndrom. *Die Kulturpflanze*, 32(1): 11-34.
- Hamner, K.C. and Bonner, J., 1938, Photoperiodism in relation to hormones as factors in floral initiation and development. *Bot. Gaz.*, 100(2): 388-431.

- Hamner, K.C., 1940, Interrelation of light and darkness in photoperiodic induction. *Bot. Gaz.*, 101(3): 658-687.
- Hänzelmann, S., Castelo, R. and Guinney, J., 2013, GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics*, 14(1): 1-15.
- Harlan, J.R., De Wet, J.M.J. and Price, E.G., 1973, Comparative evolution of cereals. *Evol.*, 27(2): 311-325.
- Harrison: J. and Dale, H.M., 1966, The effect of grazing and clipping on the control of wild carrot. *Weeds*, 14(4): 285-288.
- Hauser, T.P. and Bjørn, G.K., 2001, Hybrids between wild and cultivated carrots in Danish carrot fields. *Genet. Resour. Crop Evol.*, 48(5): 499-506.
- Hauser, T.P., 2002, Frost sensitivity of hybrids between wild and cultivated carrots. *Conserv. Genet.*, 3: 73–76.
- Hauser, T.P., Bjorn, G.K., 2001, Hybrids between wild and cultivated carrots in Danish carrot fields. *Genet. Res. Crop Evol.*, 48: 499–506.
- Hawthorn, L.R., 1952, January. Interrelations of soil moisture, nitrogen, and spacing in carrot seed production. *In Proc. Amer. Soc. Hort. Sci.*, 60: 321-326.
- Hawthorn, L.R., Toole, E.H. and Toole, V.K., 1962, Yield and viability of carrot seeds as affected by position of umbel and time of harvest. *In Proc. Amer. Soc. Hort. Sci.*, 80(4): 4.
- Hedden: and Mutasa-Gottgens, E., 2009, Gibberellin as a factor in floral regulatory networks. *J. Exp. Bot.*, 10(5): 670-674.
- Helliwell, C. A., Wood, C. C., Robertson, M., James Peacock, W. and Dennis, E.S., 2006, The *Arabidopsis* FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high - molecular - weight protein complex. *Plant J.*, 46(2): 183-192.
- Heywood, V.H., 1968a, The *Daucus carota*-gingidium complex. *Feddes Rep.*, 79: 66–68.

- Heywood, V.H., 1968b, *Daucus*. In: Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA (eds) *Flora Europaea*. Cambridge University Press, Cambridge, pp 373–375.
- Heywood, V.H., 1983. Relationships and evolution in the *Daucus carota* complex. *Israel J. Plant Sci.*, 32(2): 51-65.
- Heywood V.H., 2014, The socio-economic importance of the Apiales. *J Fac Pharm Istanbul* 44:113–130
- Hiller, L.K. and Kelly. W.C., 1979, Post-vernalization temperature effects on seed stalk elongation and flowering in carrots, *Daucus carota* L. *J. Amer. Soc. Hort. Sci.* 104: 253–257.
- Hoang, N.V., Choe, G., Zheng, Y., Fandiño, A.C.A., Sung, I., Hur, J., Kamran, M., Park, C., Kim, H., Ahn, H. and Kim, S., 2020b, Identification of conserved gene-regulatory networks that integrate environmental sensing and growth in the root cambium. *Curr. Biol.*, 30(15): 2887-2900.
- Hoang, N.V., Park, C., Kamran, M. and Lee, J.Y., 2020, Gene regulatory network guided investigations and engineering of storage root development in root crops. *Front. Plant Sci.*, 11, pp.762.
- Hölzer, M. and Marz, M., 2019, De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience*, 8(5): 39.
- Hornett, E.A. and Wheat, C.W., 2012, Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC genomics*, 13(1): 1-16.
- Huang, X., Chen, X.G. and Armbruster: A., 2016, Comparative performance of transcriptome assembly methods for non-model organisms. *BMC Genomics*, 17(1): 1-14.

- Hugo Ferney, G.B., Alexei, M. and Aigul, A., 2006, Evaluation of grain yield stability, reliability and cultivar recommendations in spring wheat (*Triticum aestivum* L.) from Kazakhstan and Siberia. *J. Cent. Eur. Agric.*, 7(4): 649-660.
- Hwang, Y., Lee, H., Lee, Y.S. and Cho, H.T., 2016, Cell wall-associated root hair specific 10, a proline-rich receptor-like kinase, is a negative modulator of *Arabidopsis* root hair growth. *J. Exp. Bot.*, 67(6): 2007-2022.
- Iorizzo, M., Senalik, D.A., Grzebelus, D., Bowman, M., Cavagnaro, P.F. and Matvienko, M., 2011, De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers and genetic diversity. *BMC Genomics* 12:389. doi:10.1186/1471-2164-12-389
- Iorizzo, M., Cavagnaro: F., Bostan, H., Zhao, Y., Zhang, J. and Simon: W., 2019, A cluster of MYB transcription factors regulates anthocyanin biosynthesis in carrot (*Daucus carota* L.) root and petiole. *Front. Plant Sci.*, 9: 1927.
- Iorizzo, M., Ellison, S., Senalik, D., Zeng, Satapoomin, Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro and Yildiz, M., 2016, A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.*, 48(6): 657-666.
- Iorizzo, M., Senalik, D.A., Ellison, S.L., Grzebelus, D., Cavagnaro: F., Allender, C., Brunet J., Spooner, D.M., Van Deynze A and Simon: W., 2013, Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (*Apiaceae*). *Am. J. Bot.*, 100: 930-938.
- Iovene, M., Cavagnaro: F., Senalik, D., Buell, C.R., Jiang, J. and Simon: W., 2011, Comparative FISH mapping of *Daucus* species (*Apiaceae* family).
- Jones, Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. and Pesseat, S., 2014, InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9): 1236-1240.
- Joseph, J.T., Poolakkalody, N.J. and Shah, J.M., 2018, Plant reference genes for development and stress response studies. *J. Biosci.*, 43(1): 173-187.

- Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ., 2016, Plant systematics: a phylogenetic approach, 4th edn. Sinauer Associates, Sunderland
- Junttila, O., 1991, Gibberellins and the regulation of shoot elongation in woody plants. In *Gibberellins* (pp. 199-210). Springer, New York, NY.
- Just, B.J., Santos, C.A., Yandell, B.S. and Simon: W., 2009, Major QTL for carrot color is positionally associated with carotenoid biosynthetic genes and interact epistemically in a domesticated wild carrot cross. *Theor. Appl. Genet.*, 119(7): 1155-1169.
- Just, B.J., Santos, C.A.F., Fonseca, M.E.N., Boiteux, L.S., Oloizia, B.B. and Simon: W., 2007, Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. *Theor. Appl. Genet.*, 114(4): 693-704.
- Just, B.J., Santos, C.A.F., Fonseca, M.E.N., Boiteux, L.S., Oloizia, B.B. and Simon: W., 2007, Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. *Theor. Appl. Genetic.*, 114(4): 693-704.
- Kanehisa, M., 2002, The KEGG database. *Novartis Found. Symp.*, 247: 91–103.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori M., 2004, The KEGG resource for deciphering the genome. *Nucl. Acids Res.*; 32(Database issue): D277–D2780. DOI: 10.1093/nar/gkh063.
- Keilwagen, J., Lehnert, H., Berner, T., Budahn, H., Nothnagel, T., Ulrich, D. and Dunemann, F., 2017, The terpene synthase gene family of carrot (*Daucus carota* L.): Identification of QTLs and candidate genes associated with terpenoid volatile compounds. *Front. Plant Sci.*, 8: 1930.
- Khanna, R., Kronmiller, B., Maszle, D.R., Coupland, G., Holm, M., Mizuno, T. and Wu, S.H., 2009, The *Arabidopsis* B-box zinc finger family. *Plant Cell*, 21(11): 3416-3420.

- Khatri: and Drăghici, S., 2005, Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, 21(18): 3587-3595.
- Khatri: and Drăghici, S., 2005, Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, 21(18): 3587-3595.
- Kim, D.H., 2020, Current understanding of flowering pathways in plants: focusing on the vernalization pathway in *Arabidopsis* and several vegetable crop plants. *Hortic. Environ. Biotechnol.*, 61(2): 209-227.
- Kim, H., Park, S.J., Kim, Y. and Nam, H.G., 2020, Subcellular localization of gigantea regulates the timing of leaf senescence and flowering in *Arabidopsis*. *Front. Plant Sci.*, 11: 1814.
- Kim, J.H., Choi, D. and Kende, H., 2003, The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. *Plant J.*, 36(1): 94-104.
- Kim, N.H., Jayakodi, M., Lee, S.C., Choi, B.S., Jang, W., Lee, J., Kim, H.H., Waminal, N.E., Lakshmanan, M., Van Nguyen, B. and Lee, Y.S., 2018, Genome and evolution of the shade - requiring medicinal herb *Panax ginseng*. *Plant Biotechnol. J.*, 16(11): 1904-1917.
- Kobayashi, Y. and Weigel, D., 2007, Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. *Genes Dev.*, 21(19): 2371-2384.
- Kodama, M., Brinch-Pedersen, H., Sharma, S., Holme, I.B., Joernsgaard, B., Dzhanfezova, T., Amby, D.B., Vieira, F.G., Liu, S. and Gilbert, M.T.P., 2018, Identification of transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. *BMC genomics*, 19(1): 1-13.
- Kogenaru, S., Yan, Q., Guo, Y. and Wang, N., 2012, RNA-seq and microarray complement each other in transcriptome profiling. *BMC genomics*, 13(1): 1-16.

- Kole, C., Quijada, Michaels, S.D., Amasino, R.M. and Osborn, T.C., 2001, Evidence for homology of flowering-time genes VFR2 from *Brassica rapa* and FLC from *Arabidopsis thaliana*. *Theor. Appl. Genet.*, 102(2-3): 425-430.
- Koornneef, M., Hanhart, C.J. and Van der Veen, J.H., 1991, A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Geneti.*, 229(1): 57-66.
- Kornberg, R.D., 2007, The molecular basis of eucaryotic transcription. *Cell Death Differ.*, 14(12): 1989-1997.
- Koul, A.K. and Hamal, I.A., 1989. Reproductive biology of wild and cultivated carrot (*Daucus carota* L.). *New Phytol.*, 112(3): 437-443.
- Koul, Sharma, N. and Koul, A.K., 1993, Pollination biology of Apiaceae. *Curr. Sci.*, pp.219-222.
- Kozera, B. and Rapacz, M., 2013, Reference genes in real-time PCR. *J. Appl. Genet.*, 54: 391– 406. PMID: 24078518
- Kulkarni C.C., Manikanta, D.S. and Cholin, S.S., 2021a. Genetic divergence analysis among indigenous temperate and tropical region adapted accessions of carrot (*Daucus carota* L.) The Pharma Innovation Journal 10(12): 2321-2326.
- Kulkarni, C.C., Cholin, S.S., Mesta, R.K., Patil, H.B., Awati, M., and Mulla, S.W. 2021b. Assessment of stability and adaptability of carrot (*Daucus carota* sub spp. *sativus*) cultivars in North Karnataka. The Pharma Innovation Journal 10(12): 632-636.
- Kulkarni, C.C., Cholin, S.S., Mesta, R.K., Patil, H.B., Awati, M., and Mulla, S.W. 2022 Genotype x environment interaction and stability analysis across environments for yield related traits of carrot (*Daucus carota* subspp. *sativus*) cultivars in North Karnataka. The Pharma Innovation Journal 11(1): 1353-1358.

- Kulkarni, C.C., Manikanta, D.S., Chaitra, A.P. and Sarvamangala, S., 2019, Character association studies by correlation estimates of diverse carrot germplasm lines evaluated under tropical climate. *J. Pharmacogn. Phytochem.*, 8(4): 1453-1458.
- Kumar: A., Reddy, K.R., Reddy, R.V.S.K., Pandravada, S.R. and Saidaiah, 2019, Stability studies among tomato genotypes for yield and processing traits. *Int. J. Chem. Stud.*, 3(3): 17-26.
- Kumar, V., Bhatt, V. and Kumar, N., 2018, Amides from plants: structures and biological importance. In *Stud. Nat. Prod. Chem. Elsevier*. 56: 287-333.
- Kust, A.F., 1972. Inheritance and differential formation of color and associated pigments in xylem and phloem of carrots, *Daucus carota*, L. PhD, University of Wisconsin
- Lacey, E.P., 1986, The genetic and environmental control of reproductive timing in a short-lived monocarpic species *Daucus carota* (Umbelliferae). *J. Ecol.*, pp.73-86.
- Lacey, E.P., 1986, The genetic and environmental control of reproductive timing in a short-lived monocarpic species *Daucus carota* (Umbelliferae). *J. Ecology*, pp.73-86.
- Lacey, E.P., 1988, Latitudinal variation in reproductive timing of a short - lived monocarp, *Daucus carota* (Apiaceae). *Ecology*, 69(1): 220-232.
- Laferriere, L., Gabelman, W. H., 1968, Inheritance of color, total carotenoids, alpha-carotene, and beta-carotene in carrots, *Daucus carota* L. *Proc. Am. Soc. Hortic. Sci.*, 93: 408–418
- Laghari, S., Kandhro, M.M., Ahmed, H.M., Sial, M.A. and Shad, M.Z., 2003, Genotype X Environment (GxE) interactions in Cotton (*Gossypium hirsutum* L.) genotypes. *Asian J. Plant Sci.*

- Lamborn, E. and Ollerton, J., 2000, Experimental assessment of the functional morphology of inflorescences of *Daucus carota* (Apiaceae): testing the fly catcher effect. *Funct. Ecol.*, 14(4): 445-454.
- Lan, T.H. and Paterson, A.H., 2000, Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. *Genetics*, 155(4): 1927-1954.
- Langmead, B. and Salzberg, S.L., 2012, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357-359.
- Langmead, B., Hansen, K.D. and Leek, J.T., 2010, Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biol.*, 11(8): 1-11.
- Le Clerc, V., Marques, S., Suel, A., Huet, S., Hamama, L., Voisine, L., Auperpin, E., Jourdan, M., Barrot, L., Prieur, R. and Briard, M., 2015, QTL mapping of carrot resistance to leaf blight with connected populations: stability across years and consequences for breeding. *Theor. Appl. Genet.*, 128(11): 2177-2187.
- Le Clerc, V., Pawelec, A., Birolleau-Touchard, C., Suel, A. and Briard, M., 2009, Genetic architecture of factors underlying partial resistance to *Alternaria* leaf blight in carrot. *Theor. Appl. Genet.*, 118(7): 1251-1259.
- Lee, Z.H., Hirakawa, T., Yamaguchi, N. and Ito, T., 2019, The roles of plant hormones and their interactions with regulatory genes in determining meristem activity. *Int. J. Mol. Sci.*, 20(16): 4065.
- Letta, T. Genotype-environment interactions and correlation among some stability parameters of yield in durum wheat (*Triticum durum* Desf) genotypes grown in South East Ethiopia. In African Crop Science Conference Proceedings 2007; 8:693-698.
- Li, B. and Dewey, C.N., 2011, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1): 1-16.

- Li, B.W., Andrews, K.W. and Pehrsson: R., 2002. Individual sugars, soluble, and insoluble dietary fiber contents of 70 high consumption foods. *J. Food Compos. Anal.*, 15(6), pp.715-723.
- Li, H. and Durbin, R., 2009, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14): 1754-1760.
- Li, L., Paolillo, D.J., Parthasarathy, M.V., DiMuzio, E.M. and Garvin, D.F., 2001, A novel gene mutation that confers abnormal patterns of β - carotene accumulation in cauliflower (*Brassica oleracea* var. botrytis). *Plant J.*, 26(1): 59-67.
- Li, S., 2015, The *Arabidopsis thaliana* TCP transcription factors: a broadening horizon beyond development. *Plant Signal. Behav.*, 10(7): e1044192.
- Liao, Y., Smyth, G.K. and Shi, W., 2014, feature Counts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30: 923–930.
- Linke, B., Alessandro, M.S., Galmarini, C.R. and Nothnagel, T., 2019, Carrot floral development and reproductive biology. In *The Carrot Genome* (pp. 27-57). Springer, Cham.
- Linnaeus, C., 1799. *Species plantarum* (Vol. 3). Impensis GC Nauk.
- Littlewood, T.D. and Evan, G.I., 1994, Transcription factors 2: helix-loop-helix. *Protein Profile*, 1(6): 635-709.
- Liu, L., Ou, C., Chen, S., Shen, Q., Liu, B., Li, M., Zhao, Z., Kong, X., Yan, X. and Zhuang, F., 2020, The response of COL and FT homologues to photoperiodic regulation in Carrot (*Daucus carota* L.). *Sci. Rep.*, 10(1): 1-13.
- Liu, R., Xu, Y.H., Jiang, S.C., Lu, K., Lu, Y.F., Feng, X.J., Wu, Z., Liang, S., Yu, Y.T., Wang, X.F. and Zhang, D.P., 2013, Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. *J. Exp. Bot.*, 64(18): 5443-5456.

- Lu, S., Van Eck, J., Zhou, X., Lopez, A.B., O'Halloran, D.M., Cosman, K.M., Conlin, B.J., Paolillo, D.J., Garvin, D.F., Vrebalov, J. and Kochian, L.V., 2006, The cauliflower gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of β -carotene accumulation. *The Plant Cell*, 18(12): 3594-3605.
- Lynch, J.P., 2011, Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. *Plant Physiol.*, 156(3): 1041-1049.
- Ma, J., Li, J., Xu, Z., Wang, F. and Xiong, A., 2018. Transcriptome profiling of genes involving in carotenoid biosynthesis and accumulation between leaf and root of carrot (*Daucus carota* L.). *Acta biochimica et biophysica Sinica*, 50(5): 481-490.
- Maass, D., Arango, J., Wüst, F., Beyer, and Welsch, R., 2009, Carotenoid crystal formation in *Arabidopsis* and carrot roots caused by increased phytoene synthase protein levels. *PloS one*, 4(7): 6373.
- Machaj, G., Bostan, H., Macko-Podgórn, A., Iorizzo, M. and Grzebelus, D., 2018, Comparative transcriptomics of root development in wild and cultivated carrots. *Genes*, 9(9): 431.
- Macko-Podgórn, A., Machaj, G., Stelmach, K., Senalik, D., Grzebelus, E., Iorizzo, M., Simon, W. and Grzebelus, D., 2017, Characterization of a genomic region under selection in cultivated carrot (*Daucus carota* subsp. *sativus*) reveals a candidate domestication gene. *Front. Plant Sci.*, 8: 12-18.
- Maksylewicz, A. and Baranski, R., 2013, Intra-population genetic diversity of cultivated carrot (*Daucus carota* L.) assessed by analysis of microsatellite markers. *Acta Biochim. Pol.*, 60(4): 58-63.
- Mandel, J.R. and Brunet, J., 2019, Gene flow in carrot. In *The carrot genome* (pp. 59-76). Springer, Cham.

- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. and Dewell, S.B., 2005, Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057): 376-380.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. and Gilad, Y., 2008, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, 18(9): 1509-1517.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. and Gilad, Y., 2008, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, 18(9): 1509-1517.
- Marques de C. L., Benda, N. D., Vaughan, M. M., Cabrera, A. R., Hung, K., Cox, T., Abdo, Z. and Allen, L. H., 2015, Teal PEA Mi-1-mediated nematode resistance in tomatoes is broken by short-term heat stress but recovers over time. *J. Nematol.*, 47: 133-140.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17(1), pp.10-12.
- Matin, M.Q.I., Rasul, M.G., Islam, A.A., Mian, M.K., Ahmed, J.U. and Amiruzzaman, M., 2017, Stability analysis for yield and yield contributing characters in hybrid maize (*Zea mays* L.). *African Journal of Agricultural Research*, 12(37):2795-2806.
- Martínez-Gómez, Crisosto, C., Bonghi, C., and Rubio, M., 2011, New approaches to prunus transcriptome analysis. *Genetica*, 139: 755–769.
- Mas, F., Harper, A., Horner, R., Welsh, T., Jaksons: and Suckling D.M., 2018, The importance of key floral bioactive compounds to honey bees for the detection and attraction of hybrid vegetable crops and increased seed yield. *J. Sci. Food Agric.*, 98(12): 4445-4453.

- McCarthy, F.M., Wang, N., Magee, G.B., Nanduri, B., Lawrence, M.L., Camon, E.B., Barrell, D.G., Hill, D.P., Dolan, M.E., Williams, W.P. and Luthe, D.S., 2006, AgBase: a functional genomics resource for agriculture. *BMC genomics*, 7(1): 1-13.
- McClung, C.R., Lou, , Hermand, V. and Kim, J.A., 2016, The importance of ambient temperature to growth and the induction of flowering. *Front. Plant Sci.*, 7: 1266.
- McGrath, K. C., Bruno, D., John M. M., Peer M. S., Cameron I. E., Donald, J. M., Wolf-Rudiger S., Michael, K., Udvardi, and Kemal K., 2005, Repressor-and activator-type ethylene response factors functioning in jasmonate signalling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol.*, 139(2): 949-959.
- Mehring-Lemper, M., 1987. Genetisch-züchterische Untersuchungen zur Schaffung von Hybridsorten bei Möhren (*Daucus carota* L.). Universität.
- Meng, G., Clausen, S.K. and Rasmussen, S.K., 2020, Transcriptome analysis reveals candidate genes related to anthocyanin biosynthesis in different carrot genotypes and tissues. *Plants*, 9(3): 344.
- Metzker, M.L., 2010, Sequencing technologies—the next generation. *Nat. Rev. Genet.*, 11(1): 31-46.
- Mezghani, N., Ruess, H., Tarchoun, N., Amor, J.B., Simon: W. and Spooner, D.M., 2018, Genotyping by sequencing reveals the origin of the Tunisian relatives of cultivated carrot (*Daucus carota*). *Genet. Resour. Crop Evol.*, 65(5): 1359-1368.
- Michaels, S.D. and Amasino, R.M., 1999, Flowering locus C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, 11(5): 949-956.
- Mohanta, T.K., Park, Y.H. and Bae, H., 2016, Novel genomic and evolutionary insight of WRKY transcription factors in plant lineage. *Sci. Rep.*, 6(1): 1-22.

- Monteiro, F., Sebastiana, M., Pais, M.S. and Figueiredo, A., 2013, Reference gene selection and validation for the early responses to downy mildew infection in susceptible and resistant *Vitis vinifera* cultivars. PloS one 8: e72998. doi: 10.1371/journal.pone.0072998 PMID: 24023800
- Morelock, T.E., TE, M. and GL, H., 1976, Glabrous seeds talk in carrot: inheritance and use as a genetic marker.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B., 2008, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, 5(7): 621-628.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B., 2008, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7): 621-628.
- Nakajima, K., Kawamura, T. and Hashimoto, T., 2006, Role of the SPIRAL1 gene family in anisotropic growth of *Arabidopsis thaliana*. *Plant cell Physiol.*, 47(4): 513-522.
- Nakajima, Y., Yamamoto, T., Muranaka, T. and Oeda, K., 1999, Genetic variation of petaloid male-sterile cytoplasm of carrots revealed by sequence-tagged sites (STSs). *Theor. Appl. Genet.*, 99(5): 837-843.
- Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N.A. and Ohtani, M., 2015, NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Front. plant Sci.*, 6: 288.
- Nicolae, M., Mangul, S., Măndoiu, I.I. and Zelikovsky A., 2011, Estimation of alternative splicing isoform frequencies from RNA - seq data. *Algorithms Mol. Biol.*, 6: 9. DOI: 10.1186/1748 - 7188 - 6 - 9
- Nieuwhof, M. and Garritsen, F.R.I.D.A., 1984, Inheritance of spine formation on seeds of carrot (*Daucus carota* L.). *Euphytica*, 33(1): 75-80.

- Nothnagel, T., Ahne, R. and Straka: , 2005, Morphology, inheritance and mapping of a compressed lamina mutant of carrot. *Plant Breed*, 124(5): 481-486.
- Nothnagel, T., Straka: and Linke, B., 2000, Male sterility in populations of *Daucus* and the development of alloplasmic male-sterile lines of carrot. *Plant Breed.*, 119: 145–152.
- O’Shea, C., Kryger, M., Stender, E.G., Kragelund, B.B., Willemoës, M. and Skriver, K., 2015, Protein intrinsic disorder in *Arabidopsis* NAC transcription factors: transcriptional activation by ANAC013 and ANAC046 and their interactions with RCD1. *Biochem. J.*, 465(2): 281-294.
- Oleszkiewicz, T., Klimek-Chodacka, M., Milewska-Hendel, A., Zubko, M., Stróż, D., Kurczyńska, E., Boba, A., Szopa, J. and Baranski, R., 2018, Unique chromoplast organisation and carotenoid gene expression in carotenoid-rich carrot callus. *Planta*, 248(6): 1455-1471.
- Oparka, K.J. and Cruz, S.S., 2000. The great escape: phloem transport and unloading of macromolecules. *Annu. Rev. Plant Biol.*, 51(1), pp.323-347.
- Osborn, T. C., Kole, C., Parkin, I. A. P., Sharpe, A. G., Kuiper, M., Lydiate, D. J. and Trick, M., 1997, Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics*, 146(3): 1123-1129.
- O'Shea, C., 2014, NAC transcription factors in stress responses and senescence: Importance of intrinsic disorder and dna binding (Doctoral dissertation, University of Copenhagen, Faculty of Science, Department of Biology)
- Oshlack, A. and Wakefield, M.J., 2009, Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct*, 4(1): 1-10.
- Oshlack, A., Robinson, M.D. and Young, M.D., 2010, From RNA-seq reads to differential expression results. *Genome Biol.*, 11(12): 1-10.
- Oshlack, A., Robinson, M.D. and Young, M.D., 2010, From RNA-seq reads to differential expression results. *Genome Biol.*, 11(12): 1-10.

- Ou, C.G., Mao, J.H., Liu, L.J., Li, C.J., Ren, H.F., Zhao, Z.W. and Zhuang, F.Y., 2017, Characterising genes associated with flowering time in carrot (*Daucus carota* L.) using transcriptome analysis. *Plant Biol.*, 19(2): 286-297.
- Paez-Garcia, A., Motes, C.M., Scheible, W.R., Chen, R., Blancaflor, E.B. and Monteros, M.J., 2015, Root traits and phenotyping strategies for plant improvement. *Plants*, 4(2): 334-355.
- Pajoro, A., Madrigal, , Muiño, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M. and Ó'Maoiléidigh, D.S., 2014, Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.*, 15(3): 1-19.
- Paparelli, E., Gonzali, S., Parlanti, S., Novi, G., Giorgi, F.M., Licausi, F., Kosmacz, M., Feil, R., Lunn, J.E., Brust, H. and van Dongen, J.T., 2012, Misexpression of a chloroplast aspartyl protease leads to severe growth defects and alters carbohydrate metabolism in *Arabidopsis*. *Plant Physiol.*, 160(3): 1237-1250.
- Parmar, D.J., Motaka, G.N., Patel, J.S. and Patel, S.G., 2016, Study on different stability procedures for yield of rice genotypes (*Oryza sativa* L.). *Int. J. Environ. Sci. Technol.*, 5(3): 1503-1514.
- Parsons, J., Matthews, W., Iorizzo, M., Roberts, and Simon, , 2015, *Meloidogyne incognita* nematode resistance QTL in carrot. *Mol. Breed.*, 35(5): 1-11.
- Patrick, J.W. and Offler, C.E., 1996. Post-sieve element transport of photoassimilates in sink regions. *J. Exp. Bot.*, 47, pp.1165-1177.
- Perrin, F., Hartmann, L., Dubois-Laurent, C., Welsch, R., Huet, S., Hamama, L., Briard, M., Peltier, D., Gagné, S. and Geoffriau, E., 2017, Carotenoid gene expression explains the difference of carotenoid accumulation in carrot root tissues. *Planta*, 245(4): 737-747.
- Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P., 2004, Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol. Lett.*, 26(6): 509-515.

- Pigliucci, M., 2005, Evolution of phenotypic plasticity: where are we going now?. *Trends Ecol. Evol.*, 20(9): 481-486.
- Poirier, V., Roumet, C. and Munson, A.D., 2018. The root of the matter: linking root traits and soil organic matter stabilization processes. *Soil Biol. Biochem.*, 120, pp.246-259.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R., 2005, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids Res.*, 33(1): 501-504.
- Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G., 1995. The constans gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*, 80(6): 847-857.
- Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens: J. and Durbin, R., 2008, A large genome center's improvements to the Illumina sequencing system. *Nat. Methods*, 5: 1005–1010.
- Que, F., Hou, X.L., Wang, G.L., Xu, Z.S., Tan, G.F., Li, T., Wang, Y.H., Khadr, A. and Xiong, A.S., 2019, Advances in research on the carrot, an important root vegetable in the Apiaceae family. *Hortic. Res.*, 6(1): 1-15.
- Que, F., Wang, G.L., Li, T., Wang, Y.H., Xu, Z.S. and Xiong, A.S., 2018, Genome-wide identification, expansion, and evolution analysis of homeobox genes and their expression profiles during root development in carrot. *Funct. Integr. Genomics*, 18(6): 685-700.
- Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K., 2003, Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol.*, 133(4): 1755-1767.

- Radonić, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W. and Nitsche, A., 2004, Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.*, 313(4): 856-862.
- Raj, R.N., Devi, C.P. and Gokulakrishnan, J., 2019, G× E interaction and stability analysis of maize hybrids using Eberhart and Russell model. *Int. J. Agric. Environ. Biotech.*, 12(1): 1-6.
- Raj, R.N., Devi, C.P. and Gokulakrishnan, J., 2019, G× E interaction and stability analysis of maize hybrids using eberhart and russell model. *Int. J. Agric. Environ. Biotechnol.*, 12(1): 1-6.
- Rau, A., Maugis-Rabusseau, C., Martin-Magniette, M.L. and Celeux, G., 2015, Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. *Bioinformatics*, 31(9): 1420-1427.
- Rédei, G.P., 1962. Supervital mutants of *Arabidopsis*. *Genetics*, 47(4): 443.
- Reeves: A., He, Y., Schmitz, R.J., Amasino, R.M., Panella, L.W. and Richards, C.M., 2007, Evolutionary conservation of the FLOWERING LOCUS C-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). *Genetics*, 176(1): 295-307.
- Reid, J.B. and Murfet, I.C., 1975, Flowering in *Pisum*: The sites and possible mechanisms of the vernalization response. *J. Exp. Bot.*, 26(6): 860-867.
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong J., Xu, C. and Merico, D., 2019, Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, cytoscape and enrichment map. *Nat. Protoc.*, 14(2): 482-517.
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong J., Xu, C. and Merico, D., 2019, Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and Enrichment Map. *Nature Protoc.*, 14(2): 482-517.

- Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J., 2007, g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.*, 35: 193-200.
- Rentel, M.C., Lecourieux, D., Ouaked, F., Usher, S.L., Petersen, L., Okamoto, H., Knight, H., Peck, S.C., Grierson, C.S., Hirt, H. and Knight, M.R., 2004, OXI1 kinase is necessary for oxidative burst-mediated signalling in *Arabidopsis*. *Nature*, 427(6977): 858-861.
- Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikat, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.J. and Xia, R., 2017, Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.*, 8(1): 1-11.
- Riechmann, J.L. and Ratcliffe, O.J., 2000, A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.*, 3: 423–434.
- Roberti, M., Polosa: L., Bruni, F., Manzari, C., Deceglie, S., Gadaleta, M.N. and Cantatore: , 2009, The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1787(5): 303-311.
- Robinson, M.D. and Oshlack, A., 2010, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, 11(3): 1-9.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1): 139-140.
- Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Pinheiro, M., Reeves: H., Putterill, J. and Coupland, G., 2001, Functional importance of conserved domains in the flowering - time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant J.*, 28(6): 619-631.

- Romeis, T., 2001, Protein kinases in the plant defence response. *Curr. Opin. Plant Biol.*, 4(5): 407-414.
- Rong, J., Janson, S., Umehara, M., Ono, M. and Vrieling, K., 2010, Historical and contemporary gene dispersal in wild carrot (*Daucus carota* ssp. *carota*) populations. *Ann. Bot.* 106: 285–296.
- Rong, J., Lammers, Y., Strasburg, J.L., Schidlo, N.S., Ariyurek, Y., De Jong, T.J., Klinkhamer, G., Smulders, M.J. and Vrieling, K., 2014, New insights into domestication of carrot from root transcriptome analyses. *BMC genomics*, 15(1): 1-15.
- Rosales-Mendoza, S. and Tello-Olea, M.A., 2015, Carrot cells: a pioneering platform for biopharmaceuticals production. *Mol. Biotechnol.*, 57(3): 219-232.
- Rosas, U., Mei, Y., Xie, Q., Banta, J.A., Zhou, R.W., Seufferheld, G., Gerard, S., Chou, L., Bhambhra, N., Parks, J.D., Flowers, J.M., McClung, C.R., Hanzawa, Y. and Purugganan, M.D., 2014, Variation in *Arabidopsis* flowering time associated with cis-regulatory variation in CONSTANS. *Nat. Commun.*, 5:3651
- Rosas, U., Mei, Y., Xie, Q., Banta, J.A., Zhou, R.W., Seufferheld, G., Gerard, S., Chou, L., Bhambhra, N., Parks, J.D. and Flowers, J.M., 2014, Variation in *Arabidopsis* flowering time associated with cis-regulatory variation in CONSTANS. *Nat. Commun.*, 5: 3651.
- Rubatsky, V. E., Quiros, C. F. and Siman: W., 1999, Carrots and related vegetable umbelliferae. CABI Publishing, NewYork, ISBN 978-0-85199-129-0.
- Ruelland, E. and Zachowski, A., 2010, How plants sense temperature. *Environ. Exp. Bot.*, 69(3): 225-232.
- Sabatini, S., Heidstra, R., Wildwater, M. and Scheres, B., 2003, SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.*, 17(3): 354-358.

- Samarth, S., Lee, R., Kelly, D., Turnbull, M., Macknight, R., Poole, A. and Jameson: , 2020, Molecular control of the floral transition in the mast seeding plant *Celmisia lyallii* (Asteraceae). *Mol. Eco.*, 30, 1846-1863.
- Santos, C. and Simon: , 2002, QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. *Mol. Genet. Genom.*, 268(1): 122-129.
- Schulz, B., Westphal, L. and Wricke, G., 1993, Linkage groups of isozymes, RFLP and RAPD markers in carrot (*Daucus carota* L. sativus). *Euphytica*, 74(1): 67-76.
- Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E., 2012, Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8): 1086-1092.
- Schneider, H.M. and Lynch, J.P., 2020. Should root plasticity be a crop breeding target. *Frontiers in Plant Science*, 11, p.546.
- Searle, I. and Coupland, G., 2004, Induction of flowering by seasonal changes in photoperiod. *The EMBO J.*, 23(6): 1217-1222.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., Amasino, R.A. and Coupland, G., 2006, The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev.*, 20(7): 898-912.
- Seboksä, G., Nigussie, M. and Bogale, G. (2001) Stability of drought tolerant maize genotypes in the drought stressed areas of Ethiopia. In Seventh Eastern and Southern Africa Regional Maize Conference; 11:301-304.
- Sharma, R.C., Morgounov, A.I., Braun, H.J., Akin, B., Keser, M., Kaya, Y., Khalikulov, Z., Ginkel, M.V., Yahyaoui, A. and Rajaram, S., 2012. Yield stability analysis of winter wheat genotypes targeted to semi-arid environments in the international winter wheat improvement program.

- Shen, X., Zhao, K., Liu, L., Zhang, K., Yuan, H., Liao, X., Wang, Q., Guo, X., Li, F. and Li, T., 2014, A role for PacMYBA in ABA-regulated anthocyanin biosynthesis in red-colored sweet cherry cv. Hong Deng (*Prunus avium* L.). *Plant Cell Physiol.*, 55(5): 862-880.
- Shendure, J. and Ji, H., 2008, Next-generation DNA sequencing *Nat. Biotechnol.*, 26(10): 1135-1145.
- Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A., 2007, The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.*, 8(9): 1-16.
- Shim, J.S. and Jang, G., 2020, Environmental signal-dependent regulation of flowering time in rice. *Int. J. Mol. Sci.*, 21(17): 6155.
- Shim, S.I. and Jørgensen, R.B., 2000, Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. *Theor. Appl. Genet.*, 101(1-2): 227-233.
- Shim, S.I. and Jørgensen, R.B., 2000, Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. *Theor. Appl. Genet.*, 101(1-2), pp.227-233.
- Shinde, G.C., Patil, J.M., Mokate, A.S., Patil, V.R. and Mehetre, S.S., 2004, AMMI analysis of cotton varieties yields trial. *J. Cotton Res. Dev.*, 18(1): 7-11.
- Shulaev, V., Korban, S.S., Sosinski, B., Abbott, A.G., Aldwinckle, H.S., Foltá, K.M., Iezzoni, A., Main, D., Arus, , Dandekar, A.M. and Lewers, K., 2008, Multiple models for Rosaceae genomics. *Plant Physiol.*, 147(3): 985-1003.
- Simão, F.A., Waterhouse, R.M., Ioannidis, , Kriventseva, E.V. and Zdobnov, E.M., 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19): 3210-3212.

- Simão, F.A., Waterhouse, R.M., Ioannidis: , Kriventseva, E.V. and Zdobnov, E.M., 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19): 3210-3212.
- Simkin, A.J., Faralli, M., Ramamoorthy, S. and Lawson, T., 2020, Photosynthesis in non - foliar tissues: implications for yield. *Plant J.*, 101(4): 1001-1015.
- Simon: W., 1996, Inheritance and expression of purple and yellow storage root color in carrot. *J. Hered.*, 87: 63–66.
- Simon, 1984, Carrot genetics. *Plant Mol. Biol. Rep.*, 2: 54–63.
- Simon: W., 1990, Carrots and other horticultural crops as a source of provitamin A carotenes. *Hort. Sci.*, 25(12): 1495-1499.
- Simon: W., 1996, Inheritance and expression of purple and yellow storage root color in carrot. *The Journal of heredity* (USA).
- Simon: W., 2000, Domestication, historical development, and modern breeding of carrot. *Plant Breed Rev.*, 19: 147–190.
- Simon: W., Freeman, R.E., Vieira, J.V., Boiteux, L.S., Briard, M., Nothnagel, T., Michalik, B. and Kwon, Y.S., 2008, Carrot. Handbook of Plant Breeding: Vegetables II: Fabaceae, Liliaceae, Solanaceae, and Umbelliferae. Edited by: Prohens J, Nuez F., New York: Springer, 327-357.
- Simon: W., Matthews, W.C. and Roberts: A., 2000, Evidence for simply inherited dominant resistance to *Meloidogyne javanica* in carrot. *Theoret. Appl. Genet.*, 100(5): 735-742.
- Simpson, G.G., 2004, The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr. Opin. Plant Biol.*, 7(5): 570-574.
- Small, E., 1978. A numerical taxonomic analysis of the *Daucus carota* complex. *Can. J. Bot.*, 56(3): 248-276.

- Song, Y.H., Ito, S. and Imaizumi, T., 2013, Flowering time regulation: photoperiod and temperature-sensing in leaves. *Trends plant Sci.*, 18(10): 575-583.
- Song, Y.H., Kubota, A., Kwon, M.S., Covington, M.F., Lee, N., Taagen, E.R., Cintrón, D.L., Hwang, D.Y., Akiyama, R., Hodge, S.K. and Huang, H., 2018, Molecular basis of flowering under natural long-day conditions in *Arabidopsis*. *Nat. Plants*, 4(10): 824-835.
- Song, Y.H., Shim, J.S., Kinmonth-Schultz, H.A. and Imaizumi, T., 2015, Photoperiodic flowering: time measurement mechanisms in leaves. *Annu. Rev. Plant Biology*, 66: 441-464.
- Souer, E., van Houwelingen, A., Kloos, D., Mol, J. and Koes, R., 1996, The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*, 85(2): 159-170.
- Soufflet-Freslon, V., Jourdan, M., Clotault, J., Huet, S., Briard, M., Peltier, D. and Geoffriau, E., 2013, Functional gene polymorphism to reveal species history: the case of the CRTISO gene in cultivated carrots. *PloS One*, 8(8): e70801.
- Spaldon, S., Samnotra, R.K., Dolkar, R. and Choudhary, D., 2017, Stability analysis and genotype $\times$ environment interaction of quality traits in tomato (*Solanum lycopersicum* L.). *Int. J. Curr. Microbiol. App. Sci.*, 6(2): 1506-1515.
- Spooner, D.M., Simon, W., Senalik, D. and Iorizzo, M., 2019, Carrot organelle genomes: organization, diversity, and inheritance. In *The Carrot Genome* (pp. 205-223). Springer, Cham.
- Spriggs, A., Henderson, S.T., Hand, M.L., Johnson, S.D., Taylor, J.M. and Koltunow, A., 2018, Assembled genomic and tissue-specific transcriptomic data resources for two genetically distinct lines of Cowpea (*Vigna unguiculata* L.). *Gates open Res.*, 2: 157-162.
- Sripathi, V.R., Anche, V.C., Gossett, Z.B. and Walker, L.T., 2021, Recent applications of RNA sequencing in food and agriculture.

- Stein, M. and Nothnagel, T.H., 1995, Some remarks on carrot breeding (*Daucus carota sativus* Hoffm.). *Plant Breeding*, 114(1): 1-11.
- Steward, F.C., Mapes, M.O. and Mears, K., 1958, Growth and organized development of cultured cells. II. Organization in cultures grown from freely suspended cell. *Am. J. Bot.*, 45(10): 705-708.
- Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Más: , Panda, S., Kreps, J.A. and Kay, S.A., 2000, Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. *Science*, 289(5480): 768-771.
- Subramanian, A., Tamayo: , Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.*, 102(43): 15545-15550.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D. and Schmidt, D., 2008, A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 321(5891), pp.956-960.
- Takada, S. and Goto, K., 2003, TERMINAL FLOWER2, an *Arabidopsis* homolog of HETEROCHROMATIN PROTEIN1, counteracts the activation of FLOWERING LOCUS T by CONSTANS in the vascular tissues of leaves to regulate flowering time. *Plant Cell*, 15(12): 2856-2865.
- Tan, J., Wang, C., Xiang, B.I.N., Han, R. and Guo, Z., 2013, Hydrogen peroxide and nitric oxide mediated cold - and dehydration - induced myo - inositol phosphate synthase that confers multiple resistances to abiotic stresses. *Plant, Cell Environ.*, 36(2): 288-299.
- Tang, D., Wang, G. and Zhou, J.M., 2017, Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell*, 29(4): 618-637.
- Tanumihardjo, S., 2012, Carotenoids and human health. Springer, New York.

- Teakle, G.R., Manfield, I.W., Graham, J.F. and Gilmartin: M., 2002, *Arabidopsis thaliana* GATA factors: organisation, expression and DNA-binding characteristics. *Plant Mol. Biol.*, 50(1): 43-56.
- Teige, M., Scheikl, E., Eulgem, T., Dóczi, R., Ichimura, K., Shinozaki, K., Dangl, J.L. and Hirt, H., 2004, The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol. Cell*, 15(1): 141-152.
- The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucl Acids Res.* 2015;43 (Database issue): D1049–D1056. DOI: 10.1093/nar/gku1179
- Theißen, R. and Melzer, 2013, Flower Development, Genetics of, Editor(s): Stanley Maloy, Kelly Hughes, Brenner's Encyclopedia of Genetics (Second Edition), Academic Press, pp.67-71.
- Thellung, A., 1926, *Daucus*. In: Hegi G (ed) *Illustrierte Flora von Mitteleuropa*. J. F. Lehmanns Verlag, München, pp 1501–1526.
- Thompson, D.J., 1960, Studies on the inheritance of male-sterility and other characters in the carrot, *Daucus carota* L. var. sativa. Cornell University.
- Tian, C., Jiang, Q., Wang, F., Wang, G.L., Xu, Z.S. and Xiong, A.S., 2015, Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. *PLoS One*, 10(2): e0117569.
- Tian, C., Jiang, Q., Wang, F., Wang, G.L., Xu, Z.S. and Xiong, A.S., 2015, Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. *PLoS One*, 10(2): e0117569.
- Tiawari, D.K., Pandey, P., Singh, R.K., Singh, S.P. and Singh, S.B. Genotype x environment interaction and stability analysis in elite clones of sugarcane (*Saccharum officinarum* L.). *International Journal of Plant Breeding and Genetics*, 2011; 5(1):93-98
- Trapnell, C. and Salzberg, S. 2009, How to map billions of short reads onto genomes. *Nat. Biotechnol.*

- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L., 2012, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Prot.*, 7(3): 562-578.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G. and Van Baren, M.J., 2010, Trans - script assembly and quantification by RNA - seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28: 511–515.
- Trujillo, M., Ichimura, K., Casais, C. and Shirasu, K., 2008, Negative regulation of PAMP-triggered immunity by an E3 ubiquitin ligase triplet in *Arabidopsis*. *Curr. Biol.*, 18(18): 1396-1401.
- Tsuji, H., 2017, Molecular function of florigen. *Breed. Sci.*, 67(4): 327-332.
- Turck, F., Fornara, F. and Coupland, G., 2008, Regulation and identity of florigen: Flowering Locus T moves center stage. *Annu. Rev. Plant Biol.*, 59: 573-594.
- Turner, S.D., Maurizio: L., Valdar, W., Yandell, B.S. and Simon: W., 2018, Dissecting the genetic architecture of shoot growth in carrot (*Daucus carota* L.) using a diallel mating design. *G3: Genes, Genomes, Genetics*, 8(2): 411-426.
- Umiel, N. and Gabelman, W. H., 1972, Inheritance of root color and carotenoid synthesis in carrot, *Daucus carota* L. Orange vs. red. *J. Am. Soc. Hort. Sci.*, 97:453–460.
- Uncu, A.O. and Uncu, A.T., 2020, High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota* L.) genome with chromosome-anchored markers. *Biotechnol. Biotechnol. Equip.*, 34(1): 1-9.
- Van Verk, M.C., Hickman, R., Pieterse, C.M. and Van Wees, S.C., 2013, RNA-Seq: revelation of the messengers. *Trends Plant Sci.*, 18(4): 175-179.
- Vavilov, N.I., Vavilov, M.I., Vavilov, N.Í. and Dorofeev, V.F., 1992, Origin and geography of cultivated plants. Cambridge University Press.

- Verkaar, H.J. and Schenkeveld, A.J., 1984, On the ecology of short-lived forbs in chalk grasslands-life-history characteristics. *New Phytol.*, 98:659–672
- Verkaar, H.J. and Schenkeveld, A.J., 1984, On the ecology of short - lived forbs in chalk grasslands: life - history characteristics. *New Phytol.*, 98(4): 659-672.
- Villeneuve, F. and Latour, F., 2014, September. Influence of sowing time and chilling exposure on flower induction in carrot (*Daucus carota* L.). In *International Symposium on Carrot and Other Apiaceae*, 1153 (pp. 47-54).
- Vivek, B.S. and Simon: W., 1999, Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. sativus). *Theor. Appl. Genet.*, 99(1-2): 58-64.
- Waines, J.G. and Ehdaie, B., 2007, Domestication and crop physiology: roots of green-revolution wheat. *Ann. Bot.*, 100(5): 991-998.
- Wang, J., Yu, J., Li, Y., Wei, C., Guo, H., Liu, Y., Zhang, J., Li, X. and Wang, X., 2020, Sequential Paleotetraploidization shaped the carrot genome. *BMC Plant Biol.*, 20(1): 1-14.
- Wang, J.W., Czech, B. and Weigel, D., 2009, miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell*, 138(4): 738-749.
- Wang, S.L., Viswanath, K.K., Tong, C.G., An, H.R., Jang, S. and Chen, F.C., 2019, Floral induction and flower development of orchids. *Front. Plant Sci.*, 10: 1258.
- Wang, Z., Gerstein, M. and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10(1): 57-63.
- Weitzel, C., Rønsted, N., Spalik, K. and Simonsen, H.T., 2014, Resurrecting deadly carrots: towards a revision of Thapsia (Apiaceae) based on phylogenetic analysis of nrITS sequences and chemical profiles. *Bot. J. Linn. Soc.*, 174(4): 620-636.

- Westmoreland, D. and Muntan, C., 1996, The influence of dark central florets on insect attraction and fruit production in Queen Anne's Lace (*Daucus carota* L.). *Am. Midl. Nat.*, pp.122-129.
- Whalley, H.J. and Knight, M.R., 2013, Calcium signatures are decoded by plants to give specific gene responses. *New Phytol.*, 197: 690-693.
- Wijnheijmer, E.H.M., Brandenburg, W.A. and Ter Borg, S.J., 1989, Interactions between wild and cultivated carrots (*Daucus carota* L.) in the Netherlands. *Euphytica*, 40(1): 147-154.
- Wils, C.R. and Kaufmann, K., 2017, Gene-regulatory networks controlling inflorescence and flower development in *Arabidopsis thaliana*. *Biochimica et Biophysica Acta (BBA)-Gene Regul. Mech.*, 1860(1): 95-105.
- Wohlfeiler, J., Alessandro, M.S., Cavagnaro, F. and Galmarini, C.R., 2019, Multiallelic digenic control of vernalization requirement in carrot (*Daucus carota* L.). *Euphytica*, 215(2): 37.
- Wu, X.J., Sun, S., Xing, G.M., Wang, G.L., Wang, F., Xu, Z.S., Tian, Y.S., Hou, X.L. and Xiong, A.S., 2017, Elevated carbon dioxide altered morphological and anatomical characteristics, ascorbic acid accumulation, and related gene expression during taproot development in carrots. *Front. Plant Sci.*, 7: 2026.
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S. and Zhou, X., 2014, SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, 30(12): 1660-1666.
- Xing, W., Liao, J., Cai, M., Xia, Q., Liu, Y., Zeng, W. and Jin, X., 2017, De novo assembly of transcriptome from *Rhododendron latoucheae* Franch. using Illumina sequencing and development of new EST-SSR markers for genetic diversity analysis in *Rhododendron*. *Tree Genet. Genomes*, 13(3): 53.
- Xu, Z.S., Tan, H.W., Wang, F., Hou, X.L. and Xiong, A.S., 2014, CarrotDB: a genomic and transcriptomic database for carrot. *Database*.

- Xu, Z.S., Yang, Q.Q., Feng, K., Yu, X. and Xiong, A.S., 2020, DcMYB113, a root - specific R2R3 - MYB, conditions anthocyanin biosynthesis and modification in carrot. *Plant Biotechnol. J.*, 18(7): 1585-1597.
- Yang, H., Liu, J., Wen, X. and Lu, C., 2015, Molecular mechanism of photosystem I assembly in oxygenic organisms. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1847(9): 838-848.
- Yao, X., Feng, H., Yu, Y., Dong, A. and Shen, W.H., 2013, SDG2-mediated H3K4 methylation is required for proper *Arabidopsis* root growth and development. *PLoS One*, 8(2): e56537.
- Yau, Y.Y. and Simon: W., 2003, A 2.5-kb insert eliminates acid soluble invertase isozyme II transcript in carrot (*Daucus carota L.*) roots, causing high sucrose accumulation. *Plant Mol. Biol.*, 53(1): 151-162.
- Yildiz, M., Willis, D.K., Cavagnaro: F., Iorizzo, M., Abak, K. and Simon: W., 2013, Expression and mapping of anthocyanin biosynthesis genes in carrot. *Theor. Appl. Genet.*, 126(7): 1689-1702.
- Yu, R., Xu, L., Zhang, W., Wang, Y., Luo, X., Wang, R., Zhu, X., Xie, Y., Karanja, B. and Liu, L., 2016, De novo taproot transcriptome sequencing and analysis of major genes involved in sucrose metabolism in radish (*Raphanus sativus L.*). *Front. Plant Sci.*, 7: 585-589.
- Yuan, H., Owsiany, K., Sheeja, T.E., Zhou, X., Rodriguez, C., Li, Y., Welsch, R., Chayut, N., Yang, Y., Thannhauser, T.W. and Parthasarathy, M.V., 2015, A single amino acid substitution in an ORANGE protein promotes carotenoid overaccumulation in *Arabidopsis*. *Plant Physiol.*, 169(1): 421-431.
- Yuan, H., Zhang, J., Nageswaran, D. and Li, L., 2015, Carotenoid metabolism and regulation in horticultural crops. *Hortic. Res.*, 2(1): 1-11.
- Zadražnik, T., Moen, A. and Šuštar-Vozlič, J., 2019, Chloroplast proteins involved in drought stress response in selected cultivars of common bean (*Phaseolus vulgaris L.*). *3 Biotech.*, 9(9): 1-15.

- Zaki, H.E.M., Takahata, Y. and Yokoi, S., 2012, Analysis of the morphological and anatomical characteristics of roots in three radish (*Raphanus sativus*) cultivars that differ in root shape. *J. Hortic. Sci. Biotechnol.*, 87(2): 172-178.
- Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R. and Simon, , 2012, Using next - generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.*, 99(2): 193-208.
- Zdobnov, E.M. and Apweiler, R., 2001, InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17(9): 847-848.
- Zerihun, J. GGE-biplot analysis of multi-environment yield trials of barley (*Hordeum vulgare* L.) genotypes in South eastern Ethiopia highlands. *International journal of plant breeding and genetics*, 2011.; 5(1): 59-75.
- Zhan, Z., Zhang C., Zhang H., Li, X., Wen, C. and Liang, Y., 2017, Molecular cloning, expression analysis, and subcellular localization of FLOWERING LOCUS T (FT) in carrot (*Daucus carota* L.). *Mol. Breed.*, 37(12), pp.1-9.
- Zhang, Z., Mao, C., Shi, Z. and Kou, X., 2017, The amino acid metabolic and carbohydrate metabolic pathway play important roles during salt-stress response in tomato. *Front. Plant Sci.*, 8: 1231.
- Zhao, J., Hartmann, H., Trumbore, S., Ziegler, W. and Zhang, Y., 2013, High temperature causes negative whole - plant carbon balance under mild drought. *New Phytol.*, 200(2): 330-339.
- Zhao, L., Nakazawa, M., Takase, T., Manabe, K., Kobayashi, M., Seki, M., Shinozaki, K. and Matsui, M., 2004, Overexpression of LSH1, a member of an uncharacterised gene family, causes enhanced light regulation of seedling development. *Plant J.*, 37(5): 694-706.
- Zhao, Y., Broholm, S.K., Wang F., Rijpkema, A.S., Lan, T., Albert, V.A., Teeri, T.H. and Elomaa: , 2020, TCP and MADS-box transcription factor networks regulate heteromorphic flower type identity in *Gerbera hybrida*. *Plant Physiol.*, 184(3): 1455-1468.

- Zhong Y., Cheng Y., Ruan, M., Ye, Q., Wang R., Yao, Z., Zhou, G., Liu, J., Yu, J. and Wan, H., 2021, High-Throughput SSR marker development and the analysis of genetic diversity in *Capsicum frutescens*. *Horticulturae*, 7(7): p. 187.
- Zhou, A., Sun, H., Dai, S., Feng, S., Zhang, J., Gong, S. and Wang, J., 2019, Identification of transcription factors involved in the regulation of flowering in *Adonis amurensis* through combined RNA-seq transcriptomics and iTRAQ proteomics. *Genes*, 10(4): 305.
- Zhou, X., Welsch, R., Yang, Y., Álvarez, D., Riediger, M., Yuan, H., Fish, T., Liu, J., Thannhauser, T.W. and Li, L., 2015, *Arabidopsis* OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci.*, 112(11): 3558-3563.
- Zhu, J., Zhang, L., Li, W., Han, S., Yang, W. and Qi, L., 2013, Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. *PloS one*, 8(1): e53196.
- Zhu, Y., Qian, W. and Hua, J., 2010, Temperature modulates plant defense responses through NB-LRR proteins. *PLoS pathogens*, 6(4): e1000844.

Appendix-I: Functional annotation of DEGs identified in reference-based root transcriptome (Functional gene sets)

Up regulated_Functional gene sets (Tropical Up regulated)	
Gene Group 1 Enrichment Score: 3.035762996231126	
ENTREZ_GENE_ID	Gene Name
108216263	probable WRKY transcription factor 17(LOC108216263)
108205774	probable WRKY transcription factor 30(LOC108205774)
108206785	probable WRKY transcription factor 30(LOC108206785)
108213109	probable WRKY transcription factor 53(LOC108213109)
108197514	probable WRKY transcription factor 53(LOC108197514)
108206893	probable WRKY transcription factor 71(LOC108206893)
108215357	probable WRKY transcription factor 15(LOC108215357)
108209169	probable WRKY transcription factor 53(LOC108209169)
108206592	WRKY transcription factor 22-like(LOC108206592)
108197802	probable WRKY transcription factor 48(LOC108197802)
108200883	probable WRKY transcription factor 48(LOC108200883)
108220707	probable WRKY transcription factor 40(LOC108220707)
108205320	probable WRKY transcription factor 31(LOC108205320)
Gene Group 2 Enrichment Score: 2.4272576928663674	
ENTREZ_GENE_ID	Gene Name
108227373	calcium-transporting ATPase 4, plasma membrane-type-like(LOC108227373)
108211663	phosphatidylinositol:ceramide inositolphosphotransferase 2-like(LOC108211663)
108214344	ABC transporter G family member 32-like(LOC108214344)
108212044	probable inactive receptor kinase At4g23740(LOC108212044)
108203999	probable methyltransferase PMT24(LOC108203999)
108205524	probable glycosyltransferase At5g03795(LOC108205524)
108219693	protein NUCLEAR FUSION DEFECTIVE 4-like(LOC108219693)
108201691	galacturonosyltransferase 8-like(LOC108201691)
108224670	uncharacterized LOC108224670(LOC108224670)
108193264	uncharacterized LOC108193264(LOC108193264)
108202454	probable LRR receptor-like serine/threonine-protein kinase At2g16250(LOC108202454)
108220838	protein QUIRKY(LOC108220838)
108193259	probable phospholipid-transporting ATPase 4(LOC108193259)
108209727	probable apyrase 6(LOC108209727)
108224664	protein YLS9(LOC108224664)
108219294	calcium-dependent protein kinase 20(LOC108219294)
108223888	tyrosine-sulfated glycopeptide receptor 1(LOC108223888)
108228098	uncharacterized LOC108228098(LOC108228098)
108210094	protein NRT1/ PTR FAMILY 8.1-like(LOC108210094)
108210476	protein NSP-INTERACTING KINASE 1-like(LOC108210476)
108215072	wall-associated receptor kinase-like 14(LOC108215072)
108197071	cytochrome P450 CYP71D313-like(LOC108197071)
108206646	probable aquaporin TIP-type(LOC108206646)
108210092	ferric reduction oxidase 2-like(LOC108210092)
108205878	ferric reduction oxidase 2-like(LOC108205878)
108226557	ribosome biogenesis ATPase RIX7-like(LOC108226557)
108206639	probable aquaporin TIP2-2(LOC108206639)
108219274	xyloglucan glycosyltransferase 4-like(LOC108219274)

108197440	ammonium transporter 2-like(LOC108197440)
108202799	oligopeptide transporter 1-like(LOC108202799)
108202798	probable sodium-coupled neutral amino acid transporter 6(LOC108202798)
108202415	protein MLP1-like(LOC108202415)
108210458	omega-3 fatty acid desaturase, chloroplastic(LOC108210458)
108225394	leucine-rich repeat receptor-like serine/threonine-protein kinase At1g17230(LOC108225394)
108202412	protein PLANT CADMIUM RESISTANCE 2-like(LOC108202412)
108222708	ABC transporter I family member 19(LOC108222708)
108225006	WAT1-related protein At5g07050-like(LOC108225006)
108199727	probable inactive purple acid phosphatase 2(LOC108199727)
108204705	uncharacterized LOC108204705(LOC108204705)
108195128	protein DETOXIFICATION 14-like(LOC108195128)
108218491	protein LYK5(LOC108218491)
108227677	serine/threonine-protein kinase CTR1(LOC108227677)
108199716	receptor protein kinase TMK1(LOC108199716)
108199715	uncharacterized LOC108199715(LOC108199715)
108218481	delta(12)-fatty-acid desaturase FAD2-like(LOC108218481)
108224226	protein NUCLEAR FUSION DEFECTIVE 4-like(LOC108224226)
108192436	peroxisomal adenine nucleotide carrier 1-like(LOC108192436)
108227288	ABC transporter G family member 22-like(LOC108227288)
108225756	probable serine/threonine-protein kinase At1g54610(LOC108225756)
108200473	uncharacterized LOC108200473(LOC108200473)
108218855	probable L-type lectin-domain containing receptor kinase S.5(LOC108218855)
108206211	rust resistance kinase Lr10-like (LOC108206211)
108206210	rust resistance kinase Lr10-like(LOC108206210)
108203144	ammonium transporter 1 member 1-like(LOC108203144)
108219229	uncharacterized protein At1g04910-like(LOC108219229)
108194334	nucleoside diphosphate kinase 1-like(LOC108194334)
108214633	mechanosensitive ion channel protein 6-like(LOC108214633)
108211566	probable E3 ubiquitin-protein ligase BAH1-like 1(LOC108211566)
108219992	protein NRT1/ PTR FAMILY 6.2-like(LOC108219992)
108195861	pyrophosphate-energized vacuolar membrane proton pump(LOC108195861)
108195094	formin-like protein 20(LOC108195094)
108215393	probable serine/threonine-protein kinase clkA (LOC108215393)
108193178	probable LRR receptor-like serine/threonine-protein kinase At4g36180(LOC108193178)
108220370	early nodulin-like protein 2(LOC108220370)
108219221	probable pectin methyltransferase QUA2(LOC108219221)
108209646	ion channel DMI1-like(LOC108209646)
108206965	probable inactive receptor kinase RLK902(LOC108206965)
108213476	uncharacterized LOC108213476(LOC108213476)
108221900	UPF0496 protein 4-like(LOC108221900)
108211176	UNC93-like protein 1(LOC108211176)
108193941	cytochrome P450 85A1(LOC108193941)
108224580	somatic embryogenesis receptor kinase 2-like(LOC108224580)
108211174	probable calcium-binding protein CML45(LOC108211174)
108193172	uncharacterized protein ycf36(LOC108193172)
108203130	RHOMBOID-like protein 10, chloroplastic(LOC108203130)

108209258	probable phospholipid-transporting ATPase 4(LOC108209258)
108210398	probable F-box protein At1g60180(LOC108210398)
108197376	cytochrome P450 76A2-like(LOC108197376)
108223801	cytochrome P450 71B26-like(LOC108223801)
108200053	calumenin-A-like(LOC108200053)
108218818	receptor-like protein kinase FERONIA(LOC108218818)
108223030	receptor-like protein kinase FERONIA(LOC108223030)
108211538	stress enhanced protein 2, chloroplastic(LOC108211538)
108203112	probable leucine-rich repeat receptor-like protein kinase At1g35710(LOC108203112)
108193533	probable glycosyltransferase At5g03795(LOC108193533)
108204639	protein NRT1/ PTR FAMILY 1.2-like(LOC108204639)
108213063	uncharacterized LOC108213063(LOC108213063)
108203102	proline-rich receptor-like protein kinase PERK9(LOC108203102)
108219182	amino acid permease 6-like(LOC108219182)
108220330	probable stress-associated endoplasmic reticulum protein(LOC108220330)
108214966	putative glycosyltransferase 7(LOC108214966)
108192368	probable pectinesterase/pectinesterase inhibitor 51(LOC108192368)
108202708	beta-1,3-galactosyltransferase GALT1-like(LOC108202708)
108210751	cytochrome P450 94A1-like(LOC108210751)
108211515	uncharacterized LOC108211515(LOC108211515)
108218026	protein DETOXIFICATION 16-like(LOC108218026)
108193892	protein ALTERED XYLOGLUCAN 4(LOC108193892)
108200018	uncharacterized protein At1g04910-like(LOC108200018)
108208827	protein NRT1/ PTR FAMILY 6.3-like(LOC108208827)
108200400	probable inactive leucine-rich repeat receptor-like protein kinase At1g66830(LOC108200400)
108196186	potassium channel SKOR-like(LOC108196186)
108215717	probable sugar phosphate/phosphate translocator At3g11320(LOC108215717)
108199629	putative calcium-transporting ATPase 13, plasma membrane-type(LOC108199629)
108219923	glucose-6-phosphate/phosphate translocator 2, chloroplastic-like(LOC108219923)
108203837	systemin receptor SR160(LOC108203837)
108208816	uncharacterized LOC108208816(LOC108208816)
108214943	cationic amino acid transporter 6, chloroplastic-like(LOC108214943)
108196555	ferric reduction oxidase 8, mitochondrial(LOC108196555)
108227578	probable aquaporin NIP5-1(LOC108227578)
108211107	uncharacterized LOC108211107(LOC108211107)
108192340	probable copper-transporting ATPase HMA5(LOC108192340)
108220297	uncharacterized LOC108220297(LOC108220297)
108205742	probable methyltransferase PMT15(LOC108205742)
108201911	ABC transporter G family member 14-like(LOC108201911)
108196924	inactive LRR receptor-like serine/threonine-protein kinase BIR2(LOC108196924)
108223734	cytochrome b561 and DOMON domain-containing protein At5g35735-like(LOC108223734)
108209561	protein trichome birefringence-like 19(LOC108209561)
108224881	probable phospholipid-transporting ATPase 4(LOC108224881)
108213007	gamma-glutamyltranspeptidase 3-like(LOC108213007)
108222963	RING-H2 finger protein ATL3-like(LOC108222963)
108200748	cell division control protein 2 homolog A-like(LOC108200748)
108210323	cytochrome b561 and DOMON domain-containing protein At4g12980(LOC108210323)

108224494	probably inactive leucine-rich repeat receptor-like protein kinase At5g06940(LOC108224494)
108200744	uncharacterized LOC108200744(LOC108200744)
108194233	CASP-like protein 4B1(LOC108194233)
108204568	ABC transporter B family member 21-like(LOC108204568)
108220651	NDR1/HIN1-Like protein 3(LOC108220651)
108216436	protein NRT1/ PTR FAMILY 6.2-like(LOC108216436)
108214900	probable galacturonosyltransferase 14(LOC108214900)
108215666	uncharacterized protein At1g04910-like(LOC108215666)
108224857	protein ABHD4(LOC108224857)
108194216	vacuolar iron transporter 1(LOC108194216)
108206086	probable serine/threonine-protein kinase Cx32, chloroplastic(LOC108206086)
108220638	probable sodium/metabolite cotransporter BASS1, chloroplastic(LOC108220638)
108214124	probable prolyl 4-hydroxylase 3(LOC108214124)
108197272	uncharacterized LOC108197272(LOC108197272)
108211441	putative ALA-interacting subunit 4(LOC108211441)
108223696	protein trichome birefringence-like(LOC108223696)
108219099	protein STRICTOSIDINE SYNTHASE-LIKE 2-like(LOC108219099)
108206842	RING-H2 finger protein ATL3-like(LOC108206842)
108224075	cytochrome P450 84A1-like(LOC108224075)
108214499	formin-like protein 1(LOC108214499)
108219859	uncharacterized LOC108219859(LOC108219859)
108219476	NDR1/HIN1-like protein 12(LOC108219476)
108209131	CSC1-like protein ERD4(LOC108209131)
108224833	transmembrane protein 184C-like(LOC108224833)
108193809	calcium-transporting ATPase 12, plasma membrane-type-like(LOC108193809)
108193807	uncharacterized LOC108193807(LOC108193807)
108227891	diacylglycerol kinase 2(LOC108227891)
108214867	probable sodium-coupled neutral amino acid transporter 6(LOC108214867)
108199929	cellulose synthase-like protein D3(LOC108199929)
108219462	leucine-rich repeat receptor-like protein kinase PXL1(LOC108219462)
108196481	uncharacterized protein At2g39920-like(LOC108196481)
108209500	ABC transporter B family member 20-like(LOC108209500)
108224819	probable folate-biopterin transporter 4(LOC108224819)
108206815	probable calcium-binding protein CML30(LOC108206815)
108223664	uncharacterized LOC108223664(LOC108223664)
108196854	protein ZINC INDUCED FACILITATOR-LIKE 1-like(LOC108196854)
108223280	uncharacterized protein At1g04910-like(LOC108223280)
108196468	probable serine/threonine-protein kinase At1g54610(LOC108196468)
108215999	cation/calcium exchanger 1(LOC108215999)
108211019	probable E3 ubiquitin-protein ligase LOG2(LOC108211019)
108215612	probable aquaporin PIP2-4(LOC108215612)
108222886	uncharacterized LOC108222886(LOC108222886)
108193777	probable receptor-like protein kinase At1g80640(LOC108193777)
108209479	leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1(LOC108209479)
108227096	protein DETOXIFICATION 42(LOC108227096)
108193390	TVP38/TMEM64 family membrane protein slr0305-like(LOC108193390)
108215221	protein kinase APK1B, chloroplastic-like(LOC108215221)

108224030	mitogen-activated protein kinase 3-like(LOC108224030)
108194155	lamin-like protein(LOC108194155)
108193005	3-ketoacyl-CoA synthase 11-like(LOC108193005)
108202575	probable receptor-like protein kinase At5g18500(LOC108202575)
108210618	glutamate receptor 3.3-like(LOC108210618)
108193766	uncharacterized LOC108193766(LOC108193766)
108227853	probable LRR receptor-like serine/threonine-protein kinase IRK(LOC108227853)
108219424	amino acid permease 3-like(LOC108219424)
108223253	calcium uniporter protein 6, mitochondrial-like(LOC108223253)
108227466	cellulose synthase-like protein D3(LOC108227466)
108209463	U-box domain-containing protein 33-like(LOC108209463)
108220950	ABC transporter B family member 4-like(LOC108220950)
108218649	protein QUIRKY(LOC108218649)
108224392	E3 ubiquitin-protein ligase ATL6(LOC108224392)
108210986	probable folate-biopterin transporter 7(LOC108210986)
108211751	purine permease 1-like(LOC108211751)
108206770	uncharacterized LOC108206770(LOC108206770)
108211748	protein IRX15-LIKE(LOC108211748)
108196807	uncharacterized LOC108196807(LOC108196807)
108209446	ABC transporter B family member 4-like(LOC108209446)
108221701	uncharacterized LOC108221701(LOC108221701)
108194120	RING-H2 finger protein ATL16-like(LOC108194120)
108216717	CASP-like protein 4C2(LOC108216717)
108210585	mechanosensitive ion channel protein 6-like(LOC108210585)
108222840	probable pectin methyltransferase QUA2(LOC108222840)
108214797	uncharacterized LOC108214797(LOC108214797)
108223606	probable LRR receptor-like serine/threonine-protein kinase At4g26540(LOC108223606)
108196792	ABC transporter G family member 21(LOC108196792)
108196408	sphingosine-1-phosphate lyase(LOC108196408)
108209804	cytochrome P450 98A2-like(LOC108209804)
108206739	oleosin 1-like(LOC108206739)
108205587	delta(8)-fatty-acid desaturase(LOC108205587)
108218609	uncharacterized LOC108218609(LOC108218609)
108219374	cationic amino acid transporter 4, vacuolar-like(LOC108219374)
108222820	respiratory burst oxidase homolog protein C-like(LOC108222820)
108202130	probable LRR receptor-like serine/threonine-protein kinase At5g10290(LOC108202130)
108223191	calcium-transporting ATPase 4, endoplasmic reticulum-type(LOC108223191)
108200976	RING-H2 finger protein ATL46(LOC108200976)
108195230	uncharacterized LOC108195230(LOC108195230)
108217440	delta(12)-fatty-acid desaturase FAD2-like(LOC108217440)
108218971	auxin transporter-like protein 3(LOC108218971)
108213991	mechanosensitive ion channel protein 6-like(LOC108213991)
108204415	uncharacterized LOC108204415(LOC108204415)
108215520	cytochrome P450 714A1-like(LOC108215520)
108215136	inactive protein kinase SELMODRAFT_444075-like(LOC108215136)
108193303	protein trichome birefringence-like 19(LOC108193303)
108197513	protein NRT1/ PTR FAMILY 5.7-like(LOC108197513)
108223557	shaggy-related protein kinase alpha-like(LOC108223557)

108219342	reticulon-like protein B21(LOC108219342)
108207849	uncharacterized LOC108207849(LOC108207849)
108223934	cyclin-dependent kinase B2-1-like(LOC108223934)
108192906	cytochrome P450 71B34-like(LOC108192906)
108202096	protein ZINC INDUCED FACILITATOR 1-like(LOC108202096)
Gene Group 3 Enrichment Score: 2.329097891155712	
ENTREZ_GENE_ID Gene Name	
108214929	probable xyloglucan endotransglucosylase/hydrolase protein 7(LOC108214929)
108228029	xyloglucan endotransglucosylase/hydrolase protein 9-like(LOC108228029)
108217608	xyloglucan endotransglucosylase/hydrolase 2-like(LOC108217608)
108215374	xyloglucan endotransglucosylase/hydrolase protein 9-like(LOC108215374)
108217616	xyloglucan endotransglucosylase/hydrolase 2-like(LOC108217616)
108200367	probable xyloglucan endotransglucosylase/hydrolase protein 27(LOC108200367)
Gene Group 4 Enrichment Score: 2.2058137725636593	
ENTREZ_GENE_ID Gene Name	
108205569	aspartic proteinase Asp1-like(LOC108205569)
108224980	aspartic proteinase nepenthesin-1-like(LOC108224980)
108210648	basic 7S globulin-like(LOC108210648)
108213406	basic 7S globulin-like(LOC108213406)
108210070	protein ASPARTIC PROTEASE IN GUARD CELL 1(LOC108210070)
108222995	protein ASPARTIC PROTEASE IN GUARD CELL 1(LOC108222995)
108197534	aspartyl protease family protein 2(LOC108197534)
108224258	basic 7S globulin-like(LOC108224258)
108197434	basic 7S globulin-like(LOC108197434)
108197384	aspartic proteinase A1-like(LOC108197384)
Gene Group 5 Enrichment Score: 2.1841382819565918	
ENTREZ_GENE_ID Gene Name	
108220055	probable galacturonosyltransferase-like 7(LOC108220055)
108197186	probable galacturonosyltransferase-like 9(LOC108197186)
108214245	probable galacturonosyltransferase-like 1(LOC108214245)
108216782	probable galacturonosyltransferase-like 1(LOC108216782)
Gene Group 6 Enrichment Score: 1.9623573430161476	
ENTREZ_GENE_ID Gene Name	
108223485	U-box domain-containing protein 44(LOC108223485)
108225391	E3 ubiquitin-protein ligase PUB23-like(LOC108225391)
108197101	U-box domain-containing protein 21-like(LOC108197101)
108224930	E3 ubiquitin-protein ligase PUB23-like(LOC108224930)
108226262	E3 ubiquitin-protein ligase PUB24-like(LOC108226262)
108192240	pre-mRNA-processing factor 19-like(LOC108192240)
108225748	U-box domain-containing protein 26(LOC108225748)
108209463	U-box domain-containing protein 33-like(LOC108209463)
108193130	U-box domain-containing protein 30-like(LOC108193130)
108213344	U-box domain-containing protein 21-like(LOC108213344)
Gene Group 7 Enrichment Score: 1.403264911812927	
ENTREZ_GENE_ID Gene Name	
108207846	probable pectinesterase/pectinesterase inhibitor 41(LOC108207846)
108215117	pectinesterase/pectinesterase inhibitor PPE8B-like(LOC108215117)
108192368	probable pectinesterase/pectinesterase inhibitor 51(LOC108192368)

108227732	pectinesterase(LOC108227732)
108193704	pectinesterase/pectinesterase inhibitor PPE8B-like(LOC108193704)
108222961	pectinesterase-like(LOC108222961)
Gene Group 8 Enrichment Score: 1.1611606564888892	
ENTREZ_GENE_ID Gene Name	
108206911	endoglucanase E1-like(LOC108206911)
108209705	acidic endochitinase SE2-like(LOC108209705)
108213112	probable beta-D-xylosidase 7(LOC108213112)
108210835	glucan endo-1,3-beta-glucosidase 14(LOC108210835)
108227783	glucan endo-1,3-beta-glucosidase 8-like(LOC108227783)
Gene Group 9 Enrichment Score: 1.0526971166910069	
ENTREZ_GENE_ID Gene Name	
108214960	L-ascorbate oxidase homolog(LOC108214960)
108199976	L-ascorbate oxidase homolog(LOC108199976)
108194461	multicopper oxidase LPR1-like(LOC108194461)
108202974	multicopper oxidase LPR1-like(LOC108202974)
Gene Group 10 Enrichment Score: 1.0473540476259606	
ENTREZ_GENE_ID Gene Name	
108223816	kinesin-like protein KIN-7G(LOC108223816)
108210490	kinesin-like protein KIF19(LOC108210490)
108204307	kinesin-like protein KCA2(LOC108204307)
108206154	kinesin-like protein KIN-7O(LOC108206154)
Gene Group 11 Enrichment Score: 1.0454559876338092	
ENTREZ_GENE_ID Gene Name	
108197721	calmodulin-like protein 11(LOC108197721)
108219294	calcium-dependent protein kinase 20(LOC108219294)
108211174	probable calcium-binding protein CML45(LOC108211174)
108195460	probable calcium-binding protein CML23(LOC108195460)
108192744	calcium-binding allergen Ole e 8-like(LOC108192744)
108200053	calumenin-A-like(LOC108200053)
108206815	probable calcium-binding protein CML30(LOC108206815)
108201870	calmodulin(LOC108201870)
108210400	probable calcium-binding protein CML41(LOC108210400)
108222820	respiratory burst oxidase homolog protein C-like(LOC108222820)
108195381	probable calcium-binding protein CML25(LOC108195381)
Gene Group 12 Enrichment Score: 0.8679299469327911	
ENTREZ_GENE_ID Gene Name	
108215975	flavin-dependent oxidoreductase FOX5-like(LOC108215975)
108199958	flavin-dependent oxidoreductase FOX2-like(LOC108199958)
108209693	cannabidiolic acid synthase-like 1(LOC108209693)
108212320	tetrahydrocannabinolic acid synthase-like(LOC108212320)
108199841	flavin-dependent oxidoreductase FOX2-like(LOC108199841)
Gene Group 13 Enrichment Score: 0.7927760076921999	
ENTREZ_GENE_ID Gene Name	
108228180	probable protein phosphatase 2C 39(LOC108228180)
108224021	probable protein phosphatase 2C 23(LOC108224021)
108200052	protein phosphatase 2C 37(LOC108200052)
108210557	probable protein phosphatase 2C 25(LOC108210557)

108205914	probable protein phosphatase 2C 25(LOC108205914)
108215250	probable protein phosphatase 2C 50(LOC108215250)
Gene Group 14 Enrichment Score: 0.6997129091486887	
ENTREZ_GENE_ID Gene Name	
108228075	probable disease resistance protein At4g33300(LOC108228075)
108197380	putative late blight resistance protein homolog R1B-16(LOC108197380)
108206201	probable disease resistance protein At1g58602(LOC108206201)
108196893	disease resistance protein RGA2-like(LOC108196893)
108218661	putative disease resistance protein At4g19050(LOC108218661)
108193163	putative disease resistance protein RGA4(LOC108193163)
Gene Group 15 Enrichment Score: 0.640344697349325	
ENTREZ_GENE_ID Gene Name	
108200182	uncharacterized LOC108200182(LOC108200182)
108202898	OBERON-like protein(LOC108202898)
108195847	1-phosphatidylinositol-3-phosphate 5-kinase FAB1B(LOC108195847)
108209850	uncharacterized LOC108209850(LOC108209850)
Gene Group 16 Enrichment Score: 0.5681419729078542	
ENTREZ_GENE_ID Gene Name	
108192993	homeobox protein knotted-1-like LET12(LOC108192993)
108199614	homeobox-leucine zipper protein HAT4-like(LOC108199614)
108197554	trihelix transcription factor GT-3b-like(LOC108197554)
108201888	ethylene-responsive transcription factor ERF109-like(LOC108201888)
108206539	homeobox-leucine zipper protein HAT4(LOC108206539)
108214195	homeobox-leucine zipper protein ATHB-12-like(LOC108214195)
Gene Group 17 Enrichment Score: 0.4354314784355198	
ENTREZ_GENE_ID Gene Name	
108225771	ethylene-responsive transcription factor ERF109-like(LOC108225771)
108225207	ethylene-responsive transcription factor 4-like(LOC108225207)
108216839	uncharacterized LOC108216839(LOC108216839)
108211806	NAC domain-containing protein 90(LOC108211806)
108213122	protein CUP-SHAPED COTYLEDON 3-like(LOC108213122)
108201888	ethylene-responsive transcription factor ERF109-like(LOC108201888)
108221909	ethylene-responsive transcription factor ERF014(LOC108221909)
108215467	scarecrow-like protein 8(LOC108215467)
108193893	ethylene-responsive transcription factor TINY-like(LOC108193893)
108225430	dehydration-responsive element-binding protein 1B-like(LOC108225430)
108211608	ethylene-responsive transcription factor CRF4-like(LOC108211608)
108220772	dehydration-responsive element-binding protein 1A-like(LOC108220772)
108225423	ethylene-responsive transcription factor ERF025-like(LOC108225423)
108192993	homeobox protein knotted-1-like LET12(LOC108192993)
108218420	DELLA protein GAI(LOC108218420)
108218599	scarecrow-like protein 21(LOC108218599)
108218926	NAC transcription factor 29-like(LOC108218926)
108225176	probable histone H2A.1(LOC108225176)
108200360	AP2/ERF and B3 domain-containing transcription factor RAV1-like(LOC108200360)
108213048	ethylene-responsive transcription factor ERF039-like(LOC108213048)
108220943	NAC domain-containing protein 72(LOC108220943)
108222305	B3 domain-containing protein Os01g0234100-like(LOC108222305)

108204772	transcription factor DIVARICATA-like(LOC108204772)
108227284	dehydration-responsive element-binding protein 1B-like(LOC108227284)
108197673	ethylene-responsive transcription factor ERF061(LOC108197673)
108194188	squamosa promoter-binding protein 1-like(LOC108194188)
Gene Group 18 Enrichment Score: 0.43159640925107395	
ENTREZ_GENE_ID Gene Name	
108209023	protein SRG1-like(LOC108209023)
108220956	probable 2-oxoglutarate-dependent dioxygenase AOP1(LOC108220956)
108225289	1-aminocyclopropane-1-carboxylate oxidase homolog 12-like(LOC108225289)
108223209	1-aminocyclopropane-1-carboxylate oxidase(LOC108223209)
108195869	linoleate 13S-lipoxygenase 2-1, chloroplastic(LOC108195869)
108207914	hyoscyamine 6-dioxygenase-like(LOC108207914)
Gene Group 19 Enrichment Score: 0.140706659776514	
ENTREZ_GENE_ID Gene Name	
108224951	protein unc-45 homolog A(LOC108224951)
108215665	inactive TPR repeat-containing thioredoxin TTL3(LOC108215665)
108204683	uncharacterized LOC108204683(LOC108204683)
108211369	uncharacterized LOC108211369(LOC108211369)
Gene Group 20 Enrichment Score: 0.03405005610245594	
ENTREZ_GENE_ID Gene Name	
108200102	LEC14B homolog(LOC108200102)
108225742	F-box/WD repeat-containing protein 7(LOC108225742)
108218447	enhancer of mRNA-decapping protein 4-like(LOC108218447)
108208777	WD repeat-containing protein 53(LOC108208777)
108192240	pre-mRNA-processing factor 19-like(LOC108192240)
108227307	coatamer subunit alpha-1-like(LOC108227307)

Down regulated_Functional gene sets - Temperate (Up regulated)	
Gene Group 1 Enrichment Score: 5.866255180635302	
ENTREZ_GENE_ID	Gene Name
108227591	psbP domain-containing protein 1, chloroplastic(LOC108227591)
108227226	oxygen-evolving enhancer protein 2, chloroplastic(LOC108227226)
108214298	oxygen-evolving enhancer protein 1, chloroplastic-like(LOC108214298)
108200075	oxygen-evolving enhancer protein 3-2, chloroplastic(LOC108200075)
108210722	oxygen-evolving enhancer protein 2-1, chloroplastic-like(LOC108210722)
108200841	thylakoid lumenal 19 kDa protein, chloroplastic-like(LOC108200841)
108196650	oxygen-evolving enhancer protein 1, chloroplastic-like(LOC108196650)
108224633	oxygen-evolving enhancer protein 1, chloroplastic(LOC108224633)
Gene Group 2 Enrichment Score: 4.2087645916458785	
ENTREZ_GENE_ID	Gene Name
108197355	chlorophyll a-b binding protein, chloroplastic-like(LOC108197355)
108202621	chlorophyll a-b binding protein CP24 10A, chloroplastic(LOC108202621)
108224860	chlorophyll a-b binding protein CP24 10A, chloroplastic-like(LOC108224860)
108204319	chlorophyll a-b binding protein, chloroplastic-like(LOC108204319)
108218758	photosystem II 22 kDa protein, chloroplastic(LOC108218758)
108196996	chlorophyll a-b binding protein CP29.2, chloroplastic(LOC108196996)
108197341	chlorophyll a-b binding protein, chloroplastic-like(LOC108197341)
108219925	chlorophyll a-b binding protein 6A, chloroplastic-like(LOC108219925)
108203547	chlorophyll a-b binding protein CP26, chloroplastic-like(LOC108203547)
108196117	chlorophyll a-b binding protein 4, chloroplastic(LOC108196117)
Gene Group 3 Enrichment Score: 1.3219614405096656	
ENTREZ_GENE_ID	Gene Name
108215302	transcription factor HBP-1b(c38)(LOC108215302)
108205500	bZIP transcription factor 53-like(LOC108205500)
108205499	bZIP transcription factor 53-like(LOC108205499)
108210779	transcription factor VIP1-like(LOC108210779)
108225041	bZIP transcription factor 53-like(LOC108225041)
108200313	bZIP transcription factor 53-like(LOC108200313)
Gene Group 4 Enrichment Score: 1.1710669603336705	
ENTREZ_GENE_ID	Gene Name
108205758	myb-related protein 308(LOC108205758)
108207943	ethylene-responsive transcription factor ERF071(LOC108207943)
108220671	ethylene-responsive transcription factor ERF071-like(LOC108220671)
108219053	ethylene-responsive transcription factor RAP2-13(LOC108219053)
108219813	uncharacterized LOC108219813(LOC108219813)
108205942	transcription factor MYB1R1-like(LOC108205942)
108193779	transcription factor MYB1R1-like(LOC108193779)
108221993	histone H4(LOC108221993)
108217718	transcription factor E2FC-like(LOC108217718)
108195193	ethylene-responsive transcription factor ERF039-like(LOC108195193)
108216471	probable histone H2A.1(LOC108216471)
108216091	transcription factor MYB48(LOC108216091)
108206684	ethylene-responsive transcription factor 4(LOC108206684)
108224825	histone H1(LOC108224825)
108214184	transcription factor MYB44(LOC108214184)

108224729	WUSCHEL-related homeobox 11(LOC108224729)
108224728	myb-related protein 330(LOC108224728)
108194137	NAC domain-containing protein 100-like(LOC108194137)
108203156	auxin-induced protein 22D-like(LOC108203156)
108210752	dehydration-responsive element-binding protein 3-like(LOC108210752)
108211036	transcription factor PCL1-like(LOC108211036)
108210656	NAC domain-containing protein 72-like(LOC108210656)
108203626	probable histone H2AXb(LOC108203626)
108226521	NAC domain-containing protein 104(LOC108226521)
108214829	dehydration-responsive element-binding protein 1A-like(LOC108214829)
108210076	AP2/ERF and B3 domain-containing transcription factor RAV1-like(LOC108210076)
108211876	protein CUP-SHAPED COTYLEDON 3-like(LOC108211876)
108222038	auxin-responsive protein IAA1-like(LOC108222038)
108209774	histone H3.2(LOC108209774)
108209393	histone H1-like(LOC108209393)
108193241	ethylene-responsive transcription factor ERF118(LOC108193241)
108211386	histone H2A.6-like(LOC108211386)
108221929	ethylene-responsive transcription factor ERF027-like(LOC108221929)
108220123	ethylene-responsive transcription factor 2-like(LOC108220123)
108226582	NAC domain-containing protein 83-like(LOC108226582)
108224587	NAC domain-containing protein 72-like(LOC108224587)
108216321	transcription factor DIVARICATA-like(LOC108216321)
108225250	transcription factor PCL1-like(LOC108225250)
108203779	histone H2AX(LOC108203779)
108200168	repressor of RNA polymerase III transcription MAF1 homolog(LOC108200168)
108222010	ethylene-responsive transcription factor TINY-like(LOC108222010)
108220390	ethylene-responsive transcription factor RAP2-7(LOC108220390)
108210225	histone H1(LOC108210225)
108216115	histone H4(LOC108216115)
Gene Group 5 Enrichment Score: 1.1256867709248126	
ENTREZ_GENE_ID Gene Name	
108220671	ethylene-responsive transcription factor ERF071-like(LOC108220671)
108224729	WUSCHEL-related homeobox 11(LOC108224729)
108193520	homeobox-leucine zipper protein HAT22-like(LOC108193520)
108217872	homeobox-leucine zipper protein HAT22-like(LOC108217872)
108199270	homeobox-leucine zipper protein HOX20-like(LOC108199270)
108219234	homeobox-leucine zipper protein ATHB-6(LOC108219234)
108226492	homeobox-leucine zipper protein HOX11(LOC108226492)
Gene Group 6 Enrichment Score: 0.6817628088768845	
ENTREZ_GENE_ID Gene Name	
108220419	haloacid dehalogenase-like hydrolase domain-containing protein At3g48420(LOC108220419)
108193026	haloacid dehalogenase-like hydrolase domain-containing protein 3(LOC108193026)
108211528	haloacid dehalogenase-like hydrolase domain-containing protein 3(LOC108211528)
108205075	sugar phosphatase YfbT-like(LOC108205075)
Gene Group 7 Enrichment Score: 0.5902379482921033	
ENTREZ_GENE_ID Gene Name	
108198980	disease resistance protein RPP13-like(LOC108198980)
108213911	putative disease resistance protein RGA4(LOC108213911)

108198565	putative disease resistance RPP13-like protein 3(LOC108198565)
108214503	putative disease resistance protein RGA4(LOC108214503)
Gene Group 8 Enrichment Score: 0.5859878597549588	
ENTREZ_GENE_ID Gene Name	
108200187	zinc finger protein CONSTANS-LIKE 16-like(LOC108200187)
108210932	zinc finger protein CONSTANS-LIKE 5-like(LOC108210932)
108226483	B-box zinc finger protein 23-like(LOC108226483)
108197100	B-box zinc finger protein 24-like(LOC108197100)
Gene Group 9 Enrichment Score: 0.2349723272389473	
ENTREZ_GENE_ID Gene Name	
108197143	scopoletin glucosyltransferase-like(LOC108197143)
108200751	crocin glucosyltransferase 3-like(LOC108200751)
108216569	UDP-glycosyltransferase 89B2-like(LOC108216569)
108220589	UDP-glycosyltransferase 73C6-like(LOC108220589)
108220477	UDP-glycosyltransferase 73C3-like(LOC108220477)
Gene Group 10 Enrichment Score: 0.17306567674557008	
ENTREZ_GENE_ID Gene Name	
108197198	U-box domain-containing protein 16(LOC108197198)
108199990	U-box domain-containing protein 10-like(LOC108199990)
108219530	U-box domain-containing protein 40-like(LOC108219530)
108222862	vacuolar protein 8(LOC108222862)
Gene Group 11 Enrichment Score: 0.15646005300549096	
ENTREZ_GENE_ID Gene Name	
108219326	delta(12)-fatty-acid desaturase FAD2-like(LOC108219326)
108192899	uncharacterized LOC108192899(LOC108192899)
108220854	inositol polyphosphate multikinase beta(LOC108220854)
108195187	uncharacterized LOC108195187(LOC108195187)
108197485	solute carrier family 25 member 44-like(LOC108197485)
108198633	delta(12)-fatty-acid desaturase FAD2-like(LOC108198633)
108213184	cysteine-rich repeat secretory protein 60-like(LOC108213184)
108215087	putative cyclic nucleotide-gated ion channel 15(LOC108215087)
108192490	ADP,ATP carrier protein 1, mitochondrial-like(LOC108192490)
108209341	uncharacterized LOC108209341(LOC108209341)
108194019	rhomboid-like protein 14, mitochondrial(LOC108194019)
108227339	phytoene synthase, chloroplastic(LOC108227339)
108211252	photosystem I subunit O(LOC108211252)
108215847	probable E3 ubiquitin-protein ligase LOG2(LOC108215847)
108199757	triose phosphate/phosphate translocator, chloroplastic(LOC108199757)
108213538	early nodulin-93-like(LOC108213538)
108223495	magnesium transporter MRS2-3-like(LOC108223495)
108220429	plastidic glucose transporter 4-like(LOC108220429)
108209322	O-glucosyltransferase rumi homolog(LOC108209322)
108196682	probable sodium-coupled neutral amino acid transporter 6(LOC108196682)
108218512	probable serine/threonine-protein kinase At1g01540(LOC108218512)
108194382	late embryogenesis abundant protein At1g64065-like(LOC108194382)
108223489	photosystem II 10 kDa polypeptide, chloroplastic-like(LOC108223489)
108205871	protein DETOXIFICATION 45, chloroplastic-like(LOC108205871)

108197060	lycopene epsilon cyclase, chloroplastic(LOC108197060)
108194370	uncharacterized LOC108194370(LOC108194370)
108195135	mitogen-activated protein kinase kinase 9(LOC108195135)
108192836	ent-kaurene oxidase, chloroplastic-like(LOC108192836)
108218494	beta-fructofuranosidase, soluble isoenzyme I-like(LOC108218494)
108224620	uncharacterized LOC108224620(LOC108224620)
108213510	bidirectional sugar transporter NEC1-like(LOC108213510)
108206612	PRA1 family protein E(LOC108206612)
108195504	pollen receptor-like kinase 3(LOC108195504)
108193205	solute carrier family 25 member 44-like(LOC108193205)
108210822	E3 ubiquitin-protein ligase RMA1H1-like(LOC108210822)
108218482	delta(12)-fatty-acid desaturase FAD2-like(LOC108218482)
108224990	cytochrome P450 78A5-like(LOC108224990)
108214265	bidirectional sugar transporter SWEET5-like(LOC108214265)
108200859	uncharacterized LOC108200859(LOC108200859)
108220009	60S ribosomal protein L18a-like protein(LOC108220009)
108199709	uncharacterized LOC108199709(LOC108199709)
108208900	outer envelope pore protein 16, chloroplastic-like(LOC108208900)
108226516	cytochrome c-type biogenesis CcmH-like mitochondrial protein(LOC108226516)
108197405	uncharacterized LOC108197405(LOC108197405)
108218464	protein WALLS ARE THIN 1-like(LOC108218464)
108200078	inorganic phosphate transporter 2-1, chloroplastic(LOC108200078)
108224972	serine/threonine-protein kinase CDL1-like(LOC108224972)
108199693	putative F-box protein At1g32420(LOC108199693)
108210415	probable serine/threonine-protein kinase At1g54610(LOC108210415)
108221905	uncharacterized LOC108221905(LOC108221905)
108217691	steroid 5-alpha-reductase DET2(LOC108217691)
108220363	mitochondrial import inner membrane translocase subunit TIM22-3-like(LOC108220363)
108224567	uncharacterized LOC108224567(LOC108224567)
108194309	uncharacterized LOC108194309(LOC108194309)
108219969	photosystem I reaction center subunit XI, chloroplastic-like(LOC108219969)
108203111	uncharacterized LOC108203111(LOC108203111)
108213069	transmembrane protein 234 homolog(LOC108213069)
108200429	probable serine/threonine-protein kinase WNK9(LOC108200429)
108209621	embryogenesis-associated protein EMB8-like(LOC108209621)
108228000	putative glycerol-3-phosphate transporter 1(LOC108228000)
108226457	uncharacterized LOC108226457(LOC108226457)
108223389	plastidic ATP/ADP-transporter-like(LOC108223389)
108200024	protein ABCI12, chloroplastic(LOC108200024)
108197340	cytochrome P450 78A5-like(LOC108197340)
108212276	protein PLANT CADMIUM RESISTANCE 10(LOC108212276)
108209977	aquaporin PIP2-1-like(LOC108209977)
108224523	cytochrome c biogenesis protein CCS1, chloroplastic(LOC108224523)
108206903	uncharacterized LOC108206903(LOC108206903)
108216092	4-coumarate--CoA ligase-like 5(LOC108216092)
108193493	delta(12)-fatty-acid desaturase FAD2-like(LOC108193493)
108204978	PGR5-like protein 1A, chloroplastic(LOC108204978)
108200381	mitochondrial uncoupling protein 3(LOC108200381)

108202677	delta(12)-fatty-acid desaturase FAD2-like(LOC108202677)
108211103	MDIS1-interacting receptor like kinase 2-like(LOC108211103)
108220294	uncharacterized LOC108220294(LOC108220294)
108220291	uncharacterized LOC108220291(LOC108220291)
108223735	ent-kaurene oxidase, chloroplastic-like(LOC108223735)
108220670	protein SPA, chloroplastic(LOC108220670)
108209177	uncharacterized LOC108209177(LOC108209177)
108219900	uncharacterized LOC108219900(LOC108219900)
108194621	uncharacterized LOC108194621(LOC108194621)
108213769	protein NUCLEAR FUSION DEFECTIVE 4(LOC108213769)
108216066	UPF0057 membrane protein At4g30650-like(LOC108216066)
108196533	probable LRR receptor-like serine/threonine-protein kinase At4g37250(LOC108196533)
108218743	delta(12) fatty acid desaturase FAD2-like(LOC108218743)
108196145	WAT1-related protein At3g30340-like(LOC108196145)
108204953	sugar carrier protein C-like(LOC108204953)
108209932	aquaporin PIP2-2-like(LOC108209932)
108212992	RING-H2 finger protein ATL78-like(LOC108212992)
108209926	inorganic phosphate transporter 1-4-like(LOC108209926)
108194223	uncharacterized protein At5g65660-like(LOC108194223)
108208009	uncharacterized LOC108208009(LOC108208009)
108222177	early nodulin-93-like(LOC108222177)
108193450	serine carboxypeptidase-like 33(LOC108193450)
108215659	uncharacterized LOC108215659(LOC108215659)
108227525	nucleobase-ascorbate transporter 3-like(LOC108227525)
108193054	ferric reduction oxidase 7, chloroplastic-like(LOC108193054)
108220629	E3 ubiquitin-protein ligase ATL23(LOC108220629)
108207987	copper transporter 6-like(LOC108207987)
108219470	ABC transporter B family member 1-like(LOC108219470)
108216787	uncharacterized LOC108216787(LOC108216787)
108199547	receptor-like protein kinase HSL1(LOC108199547)
108222909	glucomannan 4-beta-mannosyltransferase 1-like(LOC108222909)
108222906	glucomannan 4-beta-mannosyltransferase 1-like(LOC108222906)
108227501	F-box protein At3g07870-like(LOC108227501)
108220989	protein PAM68, chloroplastic(LOC108220989)
108208721	protein TORNADO 2(LOC108208721)
108215996	probable sugar phosphate/phosphate translocator At4g32390(LOC108215996)
108214080	uncharacterized protein At5g64816-like(LOC108214080)
108214079	uncharacterized protein At5g64816-like(LOC108214079)
108195306	cytochrome b561 and DOMON domain-containing protein At3g61750-like(LOC108195306)
108205258	phosphoenolpyruvate carboxylase kinase 1-like(LOC108205258)
108206022	protein NRT1/ PTR FAMILY 7.3-like(LOC108206022)
108224018	CSC1-like protein RXW8(LOC108224018)
108210994	photosystem I reaction center subunit XI, chloroplastic(LOC108210994)
108211375	RING-H2 finger protein ATL40-like(LOC108211375)
108196437	uncharacterized LOC108196437(LOC108196437)
108216351	uncharacterized LOC108216351(LOC108216351)
108204858	ferric reduction oxidase 7, chloroplastic-like(LOC108204858)
108206389	calmodulin-binding receptor-like cytoplasmic kinase 1(LOC108206389)

108222087	GDP-mannose transporter GONST4-like(LOC108222087)
108223234	uncharacterized LOC108223234(LOC108223234)
108217870	oleosin 18.2 kDa-like(LOC108217870)
108206763	probable GABA transporter 2(LOC108206763)
108200249	lipase(LOC108200249)
108202926	probable receptor-like serine/threonine-protein kinase At5g57670(LOC108202926)
108224755	uncharacterized LOC108224755(LOC108224755)
108220540	uncharacterized LOC108220540(LOC108220540)
108216708	photosystem II 10 kDa polypeptide, chloroplastic(LOC108216708)
108210194	protein CURVATURE THYLAKOID 1A, chloroplastic-like(LOC108210194)
108223982	RHOMBOID-like protein 9, chloroplastic(LOC108223982)
108219767	putative glycerol-3-phosphate transporter 1(LOC108219767)
108196775	uncharacterized LOC108196775(LOC108196775)
108226649	oleosin 5-like(LOC108226649)
108202902	uncharacterized LOC108202902(LOC108202902)
108206731	protein NRT1/ PTR FAMILY 1.1-like(LOC108206731)
108209793	uncharacterized LOC108209793(LOC108209793)
108210558	uncharacterized LOC108210558(LOC108210558)
108209790	hypothetical protein(LOC108209790)
108200588	early nodulin-like protein 1(LOC108200588)
108222802	photosystem II reaction center W protein, chloroplastic-like(LOC108222802)
108206332	uncharacterized membrane protein YuiD(LOC108206332)
108210161	uncharacterized LOC108210161(LOC108210161)
108223565	probable long-chain-alcohol O-fatty-acyltransferase 5(LOC108223565)
108202881	respiratory burst oxidase homolog protein B-like(LOC108202881)
108227391	uncharacterized LOC108227391(LOC108227391)
108207856	uncharacterized LOC108207856(LOC108207856)
108193298	cysteine-rich receptor-like protein kinase 3(LOC108193298)
108224320	mitogen-activated protein kinase kinase kinase NPK1-like(LOC108224320)
108208231	serine/threonine receptor-like kinase NFP(LOC108208231)
108211294	uncharacterized LOC108211294(LOC108211294)
108215118	uncharacterized LOC108215118(LOC108215118)
Gene Group 12 Enrichment Score: 0.08625152215481541	
ENTREZ_GENE_ID Gene Name	
108219702	ABC transporter I family member 6, chloroplastic(LOC108219702)
108225499	ABC transporter I family member 17(LOC108225499)
108219470	ABC transporter B family member 1-like(LOC108219470)
108214359	chaperone protein ClpB1(LOC108214359)
Gene Group 13 Enrichment Score: 0.05773585263791505	
ENTREZ_GENE_ID Gene Name	
108199871	alpha-xylosidase 1-like(LOC108199871)
108209304	glucan endo-1,3-beta-glucosidase 14-like(LOC108209304)
108211732	acidic endochitinase-like(LOC108211732)
108220497	alpha-L-fucosidase 1-like(LOC108220497)

**TRANSCRIPTOME PROFILING OF TEMPERATE AND TROPICAL ADAPTED
CARROT (*Daucus carota* L. subsp. *sativus*) CULTIVARS TO DECODE THE
CANDIDATE GENES RELATED TO ROOT ADAPTABILITY AND FLOWERING**

CHAITRA C. KULKARNI

2022

SARVAMANGALA S. CHOLIN

Major advisor

ABSTRACT

The present research work was conducted at the University of Horticultural Sciences, Bagalkot during 2017-2021. The objectives were to identify the candidate differentially expressed genes (DEGs) for root adaptability and flowering at vegetative and reproductive phases between tropical and temperate type cultivars of carrot (*Daucus carota* L. subsp. *sativus*). For transcriptome profiling, the matured storage root and floral primordia with leaf (FPL) tissues were selected.

Triplicate samples from each cultivar were used to generate twelve RNA-seq libraries with the raw reads of >140 M and >233 M reads for root and FPL tissues respectively. Computational analysis was performed at COH Bagalkot with the standard bioinformatics pipelines after trimming and filtering poor quality reads with Q <25, reference-based (TopHat2 v2.0.9) and *de novo* assemblies (Trinity v 2.12) were generated in root and FPL tissues separately for downstream analysis. BUSCO analysis and alignment of the libraries back to assembly confirmed the high quality of assemblies in both tissues. A total of 3544 and 2239 DEGs in root tissues were identified with 1818 and 1090 up-regulated genes in tropical and 1726 and 1149 genes up-regulated in temperate cultivars with reference-based and *de novo* assemblies respectively. Whereas, in FPL tissue, 1671 DEGs were identified based on *de novo* assembly. p-value <0.01 and log2 fold change ≥ 2.0 were considered for filtering DEGs. Gene ontology (GO) categorized DEGs into the biological process (BP), cellular components (CC), molecular functions (MF), and KEGG pathways in both tissues. Major pathways upregulated in temperate roots were specific to 'photosynthesis', 'glycolysis', 'plant hormone signal transduction, and 'carbon metabolism' that help in better source-sink translocation ability and accumulation of higher carotenoid in roots. Whereas, in tropical root tissue, 'plant-pathogen interaction', 'MAPK signaling pathways', 'phenylpropanoid biosynthesis' were upregulated indicating their better root plasticity, adaptation, and tolerance to biotic or abiotic stresses. Whereas, in FPL tissues, 'reproduction', 'cellular anatomical entity', 'intracellular anatomical structure', 'transferase activity were enriched.

Gene Regulatory Network (GRN) analysis of root transcriptome also confirmed the co-expression of photosynthesis (LSH10, PSAH, PSAN, PSAD, LHCB, PSB, *etc.*) and vascular cambium-patterning genes (TRN2, LSH10, ATG13610, AT1G56150, SP1L2, HB6, *etc.*) in association with stress-responsive genes (WRKY70, ERF-71, MKK9, MYB15, *etc.*) in the temperate and tropical cultivars. These are considered key regulatory genes that distinguish these two contrasting cultivar types of carrot for storage root development. Whereas, GRN for flowering-related genes for FPL tissue demonstrated the co-expression of key flowering-related genes (GI, CO, PhyA, AP2, FKF1, FT) in tropical and temperate cultivars with different levels of expression. The gene expression was consistent in both RNA-seq and qRT PCR. Stability analysis of elite cultivars across seasons identified stable cultivars in this study. Our report on comparative transcriptomics of economically important cultivars and tissues at vegetative and flowering phase and identification of key genes involved in vascular cambium patterning, biotic and abiotic stress, and the floral transition mechanism is the first comprehensive information in carrots globally. The identified DEGs and the huge genomic resource of the present study would be of great utility in "Genomics assisted breeding" of carrot.

qÁ. ,ÀÀ^aÀUÀ J, i. ZÉÅ½ÉÀ
¥ÀÉÀ^aÀUÀZÀPÀ

..Ǿġā Ƴæ̃ ĀREz fĀEī gUā Āl j ĒmīāPīð (f.Dgī.JEī.) «±ĀmuAīĀ ĒĀĀAīĀ āVĀ
 PĀĀCāIĀ āIāzjAīĀ CEĀĀA²PzĀVĀ (nDgīJEī2, J~J,J,JZī10, JZīC6 EvĀC)U¼
 , °AiĀEUªEĀ , °Ā²Āv/EĀm v½U¼° zĀw,A±ĀmuAīĀ (J~JJZīJ,i10, |J,J,JJZī, |J,J,JJEī,
 |J,J,Jr, J~IJZī¹r, |J,J,iC EvĀC) CEĀĀA²P zĀVĀU¼ eEvU, °ÁUE Gµª®AīĀz v½U¼°
 Mvq-Ƴæ̃PāIāvP CEĀĀA²PzĀVĀ (qSĒDgīPªĀAīīī70, EDgīJƳī71, JªīJ|P, JªīªĀAīīī©15
 EvĀC)U¼ eEvU , °-CªāIQĒ zēq|r¹z. EªU¼EĀ Ƴæ̃ĀR ªAīĀAvĒ CEĀĀA²P zĀVĀU¼AzĀ
 ƳjUt, ~ĀVz, EzĀ ±ĀRguĀª ĀĒE® ..Ǿġā CªāCīUĀV F Jggā «CĒ PĀgmīU¼EĀ Ƴæ̃Q,Āvz.
 JƳī.|J~ī. CAUĀA±P °ĒEQĀª-, ĄSAČv CEĀĀA²U¼UĀV ga¹z f.Dgī.JEī. Gµª®AīĀz āVĀIĀ
 , °Ā²Āv/EĀm v½U¼° ««z °Avz Ƴæ̃ĀR °ĒEQĀª-, ĄSAČv CEĀĀA²U¼ (fDAīīī, ¹N,
 |JZīªĀAīīīJ, J|2, JƳīPJƳī1, JƳīn), °-CªāIQĀIEĀ Ƴæ̃δ¹vĀ. DgīJEīJ-, ~ī āVĀ
 PĒDgīn |¹Dgī Jggq®E CEĀĀA²PzĀVĀU¼ CªāQ 1g āVz. , ,P āVĀ °ĒEQĀª °Avz°
 DyðPaĀV Ƴæ̃ĀR āZ v½U¼ª āVĀ CAUĀA±U¼ vĀ®ĒāvP Ƴæ̃ĀREz PĀjVāz ĒĀª agČ āVĀ
 ĒĀĀAīĀ PĀĀCāIĀ āIāzjAīĀ° M¼UĒEArgªª Ƴæ̃ĀR CEĀĀA²PzĀVĀU¼ UĀġw, Ą<P, eē<P āVĀ
 CfĀªP avq, āVĀ °Ē<Ē ƳjāvðĒAīĀ PĀAīĀð<zĒĒz agČAīĀ ēAUwPaĀV PĀgmīU¼°Ē āĒz®,
 , āUæªIā»wAīĀVz. Ƴæ̃v CzīAīĒz° UĀġw, ~Āz Ƴæ̃ĀR CEĀĀA²PzĀVĀU¼ª āVĀ Sī°vī
 fĒE/Ē<ĀPī AƳĒE®ª PĀgmīE ‘fĒE/Ē<ĀPī Ēg<Ē v½ CªāCī’AīĀ° Gvªª GƳAīĀPaĀVz.