Transcriptome and proteome analysis of the curcuminoid biosynthetic pathway in turmeric (*Curcuma longa* L.)

Thesis submitted to

University of Calicut

For the award of degree of

Doctor of Philosophy

(Biotechnology)

By

Deepa K

Under the guidance of

Dr. Sheeja TE

ICAR-INDIAN INSTITUTE OF SPICES RESEARCH
Kozhikode-673 012, Kerala, India

October 2017

CERTIFICATE

This is to certify that the thesis entitled "**Transcriptome and proteome analysis of the curcuminoid biosynthetic pathway in turmeric (*Curcuma longa* L.)**" submitted by Ms. Deepa K., to University of Calicut for the award of degree of Doctor of Philosophy in Biotechnology is the result of research work carried out by her in the Division of Crop Improvement and Biotechnology, ICAR-Indian Institute of Spices Research, Kozhikode, Kerala, India during the period January 2013 to October 2017. This thesis has not been submitted for the award of any other degree or diploma of this or any other University.

(Dr. Sheeja TE)

Signature of the Guide

Place: Kozhikode

Date :

DECLARATION

I hereby declare that the thesis entitled "**Transcriptome and proteome analysis of the curcuminoid biosynthetic pathway in turmeric (*Curcuma longa* L.)**" submitted for the award of the degree of Doctor of Philosophy in Biotechnology to University of Calicut contains the results of bonafide research work done by me at ICAR-Indian Institute of Spices Research, Kozhikode, Kerala under the guidance of Dr. Sheeja TE, Principal Scientist, Division of Crop Improvement and Biotechnology, ICAR-Indian Institute of Spices Research. This thesis has not been submitted for the award of any other degree or diploma of this or any other University.

(Deepa K)

Place: Kozhikode

Date:

Acknowledgement

First and foremost, I praise ***God,*** *the almighty for His showers of blessings throughout my research work to complete the research successfully.*

I take this opportunity to express my heartfelt gratitude and indebtedness to my beloved guide ***Dr. Sheeja TE****, Principal Scientist, Crop Improvement and Biotechnology Division, ICAR-Indian Institute of Spices Research (IISR), Kozhikode for her continuous support, insightful decision, thoughtful guidance, critical comments and motivation.*

My sincere thanks to ***Dr. Nirmal Babu****, Director and* ***Dr. M. Anandaraj****, former Director, ICAR-IISR for permitting and extending me the facilities to carry out this work at the ICAR-IISR, Kozhikode.*

I would like to express my heartfelt thanks to ***Dr. B. Sasikumar****, Head of Division and* ***Shri. B. Krishnamoorthy****, former Head (i/c) of Division for their constant support during my research work.*

I express my deep gratitude to ***doctoral committee members*** *for their valuable suggestions.*

I would like to thank ***Distributed Information Sub-Centre (DISC), library, PME, HRD, Administrative section****, ICAR-IISR for their help during the period of work. I owe my deep sense of gratitude to all the* ***Scientists*** *of Division of Crop Improvement and Biotechnology, Division of Crop Protection, Division of Crop Production and Post Harvest Technology for their help and encouragement.*

I thank ***AICRP on spices*** *for assisting me in turmeric field trials at Tamil Nadu Agricultural University, Coimbatore.*

It is a great pleasure to thank all my ***labmates*** *for their constant encouragement and mental support for the successful completion of this Ph. D work.*

I express my special thanks to ***supporting staff*** *of ICAR-IISR for their genuine support throughout this research work.*

I thank ***Department of Biotechnology*** *for the financial support and* ***ICAR*** *for permitting me to register for Ph. D. I thank* ***University of Calicut*** *for all the necessary help and support right from registration to completion of this Ph. D.*

Last but not the least, I would like to thank ***my daughter, parents, brother*** *and* ***husband*** *for providing me unfailing support and continuous encouragement throughout my years of study and through the process of writing the thesis. This accomplishment would not have been possible without them. Finally, my thanks go to all the people who have supported me to complete the research work.*

Deepa K

CONTENTS

Chapter	Topic	Page No.
1.	Introduction	1
2.	Review of Literature	5
3.	Methodology	31
4.	Results	62
5.	Discussion	122
6.	Summary and future aspects	148
7.	References	154
8.	Appendix	171
9.	Publications and awards	173

List of Tables

No.	Title	Page No.
1	Spectrophotometric analysis of yield and purity of RNA isolated from rhizomes and leaves using different protocols	65
2	Evaluation of yield and quality with different concentration of components in RNA isolation	67
3	Quality and yield of RNA extracted using optimized protocol from different tissues of turmeric and related spp.	69
4	Functional annotation of unigenes derived from normalized turmeric cDNA library	79
5	Statistics of putative genes involved in curcuminoid biosynthesis	80
6a	List of polyketide synthase genes identified from *C. aromatica*	81
6b	List of polyketide synthase genes identified from *C. longa*	82
7	List of differentially expressed polyketide synthase genes	83
8	Gene ontology analysis of up-regulated polyketide synthase transcripts in *C. longa*	83
9	Characteristics of calibration curves in the amplification of reference genes	86
10	Stability of reference gene expression in developing turmeric tissues	86
11	Pearson's correlation of the curcumin content with the gene expression when different tissues were analyzed at three developmental stages from two turmeric genotypes planted in three environmental conditions	97
12	Diversity of amino acid residues in the cyclization pocket, substrate binding pocket and geometry shapers of PKSs	104
13	Characteristics of substrates docked with ClPKS11 and CURS1 of *C. longa*	112
14	Details of amino acid residues involved in ClPKS11-ligand interactions	112
15a	List of putative *cis*-acting regulating elements in the sense (positive) strand of upstream sequence of *ClPKS11*	116
15b	List of putative *cis*-acting regulating elements in the anti-sense (negative) strand of upstream sequence of *ClPKS11*	117
16	List of probable MYBs binding to the upstream sequence of *ClPKS11*	117

List of Figures

No	Title	Page No.
1.	(A) Habit of *Curcuma longa* (turmeric), (B) turmeric flower, (C) turmeric rhizomes and (D) turmeric powder	7
2.	Chemical structure of curcuminoids	8
3.	Schematic representation of curcuminoid biosynthesis in turmeric	29
4.	View of turmeric seedlings from (A) Kozhikode-field, (B) Coimbatore-field, (C) Kozhikode-green house, (D) Rhizomes of IISR Prathibha and (E) Collection no. 449; (F) Cross-section of primary rhizomes from three low curcumin turmeric collections, three high curcumin turmeric varieties and three zero curcumin genotypes of *Curcuma*	35
5.	Map of different DNA rulers and protein molecular weight marker	36
6.	Map of pGEM-T Easy vector	36
7.	Schematic diagram of ds cDNA synthesis	45
8.	Outline of Duplex Specific Nuclease normalization	48
9.	Schematic outline of SMARTer ds cDNA-inverse PCR approach for full length cDNA amplification	53
10.	Agarose gel electrophoresis of RNA isolated from (a) leaves and (b) rhizomes using different protocols	64
11.	Optimization of major components in RNA isolation	66
12.	Denaturing agarose gel electrophoresis of RNA isolated from leaves (1), pseudostem (2), rhizomes (3) and roots (4)	70
13.	Electropherogram of total RNA isolated from different tissues of turmeric generated by Agilent 2100 Bioanalyzer	70
14.	Denaturing polyacrylamide gel electrophoresis of small RNA isolated from leaves (1), pseudostem (2), rhizome (3) and root (4)	71
15.	Agarose gel electrophoresis of RNA isolated from rhizomes of *C. longa* Collection no. 19 (1), *C. longa* RH 17 (2), *C. longa* Megha turmeric (3), *C. xanthorrhiza* (4), *C. amada* (5), *C. caesia* (6) and *Zingiber officinale* (7)	71
16.	PCR amplification of actin using the cDNA from rhizome	73
17.	Quality check of cDNA by qRT-PCR	73
18.	Amplification plot and melt curve of amplicon generated using mature miRNA sequences from total RNA of rhizome (1) and root (2) during qRT-PCR	75
19.	Agarose gel electrophoresis of PCR product when first strand cDNA was subjected to 18, 21, 24 and 27 number of PCR cycles	75
20.	Agarose gel electrophoresis of PCR product when the double stranded cDNA was subjected to 7, 9, 11 & 13 number of PCR cycles	76
21.	Agarose gel electrophoresis of double stranded cDNA after treating with 0.25 U (N3), 0.5 U (N2) and 1 U (N1) of DSN	76

22.	Amplification plot generated from control ds cDNA (1) and normalized ds cDNA (2) using actin primers	77
23.	Representative plates showing blue-white screening of transformants (A) Control plate (B) Experimental plate	77
24.	Representative agarose gel of PCR products of recombinant plasmids with M13 primers	78
25.	Distribution of different transcription factors in the assembled transcripts of *C. longa* and *C. aromatica*	85
26.	Agarose gel electrophoresis of real time PCR products of candidate reference genes	87
27.	Melt curve analysis of reference genes	87
28.	Average standard deviation of candidate reference genes as determined by Delta Ct method	88
29.	Average expression stability values of candidate reference genes as determined by NormFinder	88
30.	Average expression stability values of candidate reference genes as determined by GeNorm	89
31.	Average standard deviation of candidate reference genes as determined by Bestkeeper	89
32.	Relative gene expression analysis of selected genes in leaves of turmeric harvested at 60, 120 and 180 DAP	92
33.	Relative gene expression analysis of selected genes in rhizomes of turmeric harvested at 60, 120 and 180 DAP	93
34.	Comparison of relative gene expression analysis of selected genes in rhizomes vs leaves of turmeric harvested at 60, 120 and 180 DAP	94
35.	Comparison of relative gene expression analysis of selected genes in rhizomes of IISR Prathibha vs Collection no. 449 harvested from three environmental conditions	95
36.	Relative gene expression analysis of selected genes in rhizomes of IISR Prathibha harvested from three environmental conditions	96
37.	Amplification of full length ORF of reported PKSs from genomic DNA	99
38.	Schematic representation of introns in (A) *DCS* (B) *CURS1* and (C) *CURS3*	99
39.	Agarose gel showing the inverse PCR product of ~1.1 kb amplified with CURS3 IP primers from circularized cDNA template but not with non-circularized cDNA	101
40.	Comparison of 3' UTR regions of *CURS3* in *C. longa* and *C. aromatica*	101
41.	Agarose gel showing the inverse PCR product of ~1.6 kb amplified with ClPKS11 IP primers from circularized cDNA template	102
42.	Agarose gel showing the PCR product amplified with ClPKS11-FL primers from cDNA and gDNA	102

43.	Nucleotide and deduced amino acid sequence of ClPKS11	103
44.	Multiple sequence alignment of ClPKS11 with other type III PKSs	105
45.	Phylogenetic tree constructed with ClPKS11 and other type III PKSs	107
46.	Secondary structure of ClPKS11	108
47.	Predicted three dimensional structure of ClPKS11	109
48.	Ramachandran plot analysis of ClPKS11	109
49.	Interaction of substrates (a) feruloyl CoA and (b) coumaroyl CoA (c) feruloyldiketide CoA and (d) coumaroyldiketide CoA with the amino acids of ClPKS11	110-111
50.	Binding of feruloyl CoA to the active site of ClPKS11	113
51.	Alignment of *ClPKS11* intron sequences from *Curcuma* spp.	113
52.	Agarose gel showing PCR amplification of mature mRNA of *ClPKS11* with exon1-exon2 primers in cDNA of IISR Prathibha and collection no. 200	115
53.	Identification of restriction enzyme sites in the ORF of *ClPKS11*	115
54.	(a) Agarose gel showing the PCR product of gDNA-inverse PCR with PMTR-IPb primers (b) Upstream sequence of *ClPKS11* (c) Confirmation of promoter sequence using PCR amplification with PMTR primers	120
55.	SDS-PAGE analysis of protein extracted from rhizomes of (1) IISR Prathibha and (2) Collection no. 449	120
56.	2DE of protein extracted from rhizomes of IISR Prathibha (a) and collection no. 449 (b) separated by IEF	121

ABBREVIATIONS

µA	-	Microampere
µg	-	Microgram
µl	-	Microlitre
µM	-	Micromolar
ºC	-	Degree Celcius
2DE	-	2-Dimensional electrophoresis
2-PS	-	2-Pyrone synthase
4CL	-	4-Coumarate CoA: ligase
APS	-	Ammonium persulfate
BLAST	-	Basic local alignment search tool
bp	-	base pair
BSA	-	Bovine serum albumin
C3H	-	*p*-Coumarate 3-hydroxylase
C4H	-	Cinnamate 4-hydroxylase
CaCl2	-	Calcium chloride
Cbe-F	-	Coimbatore-field
cDNA	-	Complementary DNA
CHS	-	Chalcone synthase
ClPKS	-	*Curcuma longa* polyketide synthase
cm	-	Centimetre
COMT	-	Caffeoyl CoA O-methyltransferase
Cq	-	Quantification cycle
CTAB	-	Cetyltrimethylammonium bromide
CURS	-	Curcumin synthase
CYP	-	Cytochrome P450 monooxygenase
DAP	-	Days After Planting
DCS	-	Diketide CoA synthase
DEPC	-	Diethylpyrocarbonate
DNA	-	Deoxyribonucleic acid
DNase	-	Deoxyribonuclease
dNTP	-	Deoxynucleotide triphosphate
ds	-	Double stranded
DSN	-	Duplex specific nuclease
DTT	-	Dithiothreitol
EDTA	-	Ethylenediaminetetraacetic acid
EF1α	-	Elongation factor 1α
EST	-	Expressed sequence tag
g/l	-	gram per litre
Gb	-	Gigabyte
gDNA	-	Genomic DNA
H	-	Hour
HCl	-	Hydrochloric acid

HCT	-	Hydroxycinnamoyl CoA transferase
HPLC	-	High-performance liquid chromatography
IISR	-	Indian Institute of Spices Research
IPG	-	Immobilized pH gradient
IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
iTRAQ	-	Isobaric tags for relative and absolute quantitation
kb	-	Kilobase
kDa	-	Kilodalton
Koz-F	-	Kozhikode-field
Koz-GH	-	Kozhikode-green house
kvh	-	Kilovolt hour
LB	-	Luria-Bertani
LiCl	-	Lithium chloride
M	-	Molar
mg	-	Milligram
MgCl2	-	Magnesium chloride
min	-	Minute
miRNA	-	MicroRNA
ml	-	Millilitre
mM	-	Millimolar
mm	-	millimeter
mRNA	-	Messenger RNA
NaCl	-	Sodium chloride
NaOH	-	Sodium hydroxide
NCBI	-	National Center for Biotechnology Information
ng	-	Nanogram
nm	-	Nanometre
nt	-	Nucleotide
OD	-	Optical density
OKS	-	Octaketide synthase
ORF	-	Open reading frame
PAGE	-	Polyacrylamide gel electrophoresis
PAL	-	Phenylalanine ammonia lyase
PCR	-	Polymerase chain reaction
PCS	-	Pentaketide chromone synthase
PEG	-	Polyethylene glycol
pH	-	Potential of hydrogen
PKS	-	Polyketide synthase
PMTR	-	Promoter
PVP	-	Polyvinylpyrrolidone
qRT-PCR	-	Quantitative real time PCR
RACE	-	Random amplification of cDNA ends
RIN	-	RNA integrity number
RNA	-	Ribonucleic acid

RNase	-	Ribonuclease
rpm	-	Revolutions per minute
rRNA	-	Ribosomal RNA
RT	-	Room temperature
RT-PCR	-	Revrese transcription PCR
s	-	Second
SBL	-	Sequencing by ligation
SBS	-	Sequencing by synthesis
SDS	-	Sodium dodecyl sulphate
SMART	-	Switching mechanism at the 5' end of the RNA transcript
STS	-	Stibene synthase
TAL	-	Tyrosine ammonia lyase
TEMED	-	N,N,N′,N′-Tetramethylethylenediamine
Tris	-	Tris(hydroxymethyl)aminomethane
U	-	Unit
UFGT	-	UDP-glucose: flavonoid 3-O-glucosyltransferase
UGT	-	Uridine diphosphate glycosyltransferases
uORF	-	Upstream open reading frame
UTR	-	Untranslated region
UV	-	Ultraviolet
V	-	Volt
X-GAL	-	5-Bromo-4-Chloro-3-Indolyl -D-Galactopyranoside

Dedicated to My Family...

Introduction...

Introduction

Turmeric, *Curcuma longa* L. (syn. *Curcuma domestica* Valeton) is a rhizomatous perennial herb of the family Zingiberaceae. It is considered to have originated in the Indo-Malayan region (Purseglove, 1968) and has been widely distributed in the tropics of Asia, Africa and Australia. It is cultivated vegetatively as an annual crop for harvesting rhizomes, the medicinal part of the plant. It is mentioned in 'Atharva Veda', a holy treatise of the Hindus as 'Haldi' or 'Haridrar' (Shah, 1997). The dried underground rhizome of *C. longa* has been used from time immemorial as a natural dye, cosmetic, culinary ingredient, as sacred ingredient in religious functions and also used to treat various ailments in treatment systems of Ayurveda, Unani and Siddha (Velayudhan et al., 1996; Ammon and Wahl, 1991; Eigner and Scholz, 1999). India is the top producer and exporter of turmeric in the world (dasd.gov.in).

Turmeric contains proteins (6.3%), fats (5.1%), minerals (3.5%), carbohydrates (69.4%) and moisture (13.1%). The rhizomes contain 3–15% curcuminoids, which is a mixture of curcumin (diferuloylmethane), demethoxycurcumin and bisdemethoxycurcumin (Lee et al., 2016). Most of the therapeutic properties of turmeric are attributed to curcumin ($C_{21}H_{20}O_6$), a hydrophobic polyphenol which is a secondary metabolite. Curcumin was found to have multiple targets and mechanism of action, and hence is effective in treating a range of diseases including diabetes, cancer, arthritis, Alzheimer's and cardiovascular diseases (Fadus et al., 2016; Ghosh et al., 2015; Prasad et al., 2014). Indian turmeric is having high economic value due to its high curcumin content (Sasikumar, 2005). Although curcumin has been reported from various *Curcuma* spp like *C. xanthorrhiza* (1.5%), *C. angustifolia* (0.2%), *C. zedoaria* (2.0%); *C. longa* is the main source of curcumin with curcumin content ranging from 3-15% (Lee et al., 2016). High curcumin content varieties are especially sought by pharmaceutical industries. However the curcumin content was reported to vary within different accessions of *C. longa* (Wang et al., 1999; Jayaprakasha et al., 2002) and even when the same variety was planted in different agro-climatic conditions (Anandaraj et al., 2014). The exact reasons or factors affecting curcumin biosynthesis are still unknown.

After oral administration, curcumin undergoes rapid degradation at physiological pH. However, in *ex-vitro* experiments, stability of curcumin increases up to 2 hours in the

presence of serum or with cultured cells due to proteins that increases its stability (Wang et al., 1997). Hence when cultured cells incubated with curcumin after typical 4-8 h was analyzed, curcumin could have degraded and thus impossible to interpret whether the activity is due to curcumin or its degraded products. Four pathways have been described for degradation of curcumin namely, reduction, conjugation, oxidation and cleavage (Schneider et al., 2015). Reduction of curcumin results in the formation of di-, tetra-, hexa- and octahydrocurcumin. Alcohol dehydrogenase reduces curcumin to tetra- and hexahydrocurcumin in liver and an unidentified microsomal enzyme is involved in the formation of di- and octahydrocurcumin (Ireson et al., 2002). Reduction can also occur in gut by *CurA*, an NADH dependent curcumin/dihydrocurcumin reductase identified from intestinal *E. coli* (Hassaninasab et al., 2011). In plasma, curcumin and its metabolites exists as conjugates with glucuronic acid and sulphate (Pan et al., 1999). The presence of peroxidases can enhance the enzymatic oxidation of curcumin and thus decrease the stability of curcumin (Griesser et al., 2011) whereas glucuronidation enhances the stability of curcumin (Schneider et al., 2015). The prolonged incubation of curcumin in buffer resulted in the cleavage of curcumin to vanillin. Among the curcuminoids, bisdemethoxycurcumin has higher stability than demethoxycurcumin and, curcumin the least (Schneider et al., 2015). Both bisdemethoxycurcumin and demethoxycurcumin were reported to protect the curcumin from oxidation (Gordon et al., 2015). The complete oxidation of curcumin results in the formation of bicyclopentadione as the major product and many minor products through a series of reactions with a number of intermediates, each having distinct cellular targets and may amount to the "polypharmacology" of curcumin (Schneider et al., 2015).

Although numerous reports on different mechanisms of action of curcumin are available, studies related to curcumin biosynthesis and its regulation is very scarce. A precise knowledge of both structural and regulatory genes involved in curcumin biosynthetic pathway is essential to increase the curcumin content of turmeric by genetic manipulation. A basic framework of curcumin biosynthetic pathway is available based on radiotracer, molecular and enzyme kinetic studies (Roughley and Whiting, 1973; Ramirez-Ahmuda et al., 2006; Neema, 2005) and they reported that curcumin is synthesized through phenylpropanoid pathway. Phenylpropanoid pathway is well characterized in many model and non-model plants, where phenylalanine is converted to *p*-coumaroyl CoA by the action of three enzymes, phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H)

and 4-coumarate CoA: ligase (4CL). Multiple isoforms are reported for genes in phenylpropanoid pathway (Bhuiyan et al., 2009; Lillo et al., 2008). Further, based on the action of rate-limiting enzymes, the phenylpropanoid derivatives are channelized into various secondary metabolites like lignin, flavonoid etc. (Baxter and Stewart Jr, 2013). Schroder (1997) suggested that type III polyketide Katsuyama et al. (2009 a, b) synthases are involved in the biosynthesis of diarylheptanoids like curcumin. identified four polyketide synthases from turmeric namely diketide CoA synthase (DCS) and curcumin synthases (CURS 1-3) from turmeric. Apart from the structural genes of the pathway, transcription factors like MYB, WRKY, bHLH are also reported to have significant regulatory roles in secondary metabolite biosynthesis (Shi and Xie, 2014; Xie et al., 2014). Koo et al. (2013) suggested that MYB transcription factors might have a regulatory role in curcumin biosynthesis.

However, detailed studies involving identification of specific isoform and rate-limiting enzyme, gene expression analysis correlated with curcumin content and regulation of curcumin biosynthesis are yet to be carried out.

The exploitation of next generation sequencing in mining of genes related to secondary metabolite biosynthesis based on transcriptome approach has been reported (Vaidya et al., 2013; Kalra et al., 2013; Upadhyay et al., 2014). Transcriptome mining and gene expression analysis correlated to metabolite content was applied in many crops to identify the genes related to a specific pathway (Fridman et al., 2005; Roslan et al., 2012). The structural characterization of selected genes by full length cDNA amplification, sequence analysis and molecular docking provided an insight into the physiological roles of genes (Bhat et al., 2013). The direct measurement of protein expression is also essential to correlate the mRNA and protein levels in a biological system (Futcher et al., 1999). Two dimensional polyacrylamide gel electrophoresis combined with mass spectroscopic analysis was reported to identify the differentially expressed peptides (Raharjo et al., 2004; Sebastiana et al., 2013).

Aim and objectives

In spite of the immense importance of curcumin, very few studies have been conducted on the molecular mechanisms regulating its biosynthesis. Besides, it is evident from recent studies that curcumin content is highly dependent on environmental and

growth conditions. The current technological advancements regarding transcriptional analysis of turmeric genome *vis-a-vis* curcumin synthesis is a totally unexplored subject and a wealth of information can be retrieved using the state of art NGS and molecular techniques.

Thus this study was formulated with following objectives:

1. To elucidate the molecular regulatory mechanism behind curcumin biosynthesis in *C. longa* by transcriptome analysis
2. To understand the differential gene expression pattern *vis-a-vis* curcumin content under specific developmental, environmental and growth conditions
3. To study the differentially expressed proteins in correlation with biosynthesis of curcumin through comparative proteome analysis

Review of literature...

Review of literature

2.1 *Curcuma longa*

Curcuma longa L. (syn. *C. domestica* Val.) is a perennial rhizomatous erect herb with aerial parts comprising of green coloured leaves, inflorescence and pseudostem. At the base, the pseudostem is modified into mother rhizome which branches into primary, secondary and tertiary rhizomes (Sasikumar, 2005). It belongs to the Zingiberaceae family and is reported to have originated from Indo-Malayan region (Purseglove, 1968). It is distributed throughout the tropics of Asia, Africa and Australia. *C. longa* is commonly known as turmeric and it mainly refers to its rhizomes (Sasikumar, 2005) (**Fig. 1**).

The use of turmeric was mentioned in "Atharva Veda", a holy book of Hindus as a remedy to cure skin diseases, jaundice and greying of hairs. It is being used as a spice, cosmetic, food preservative and in traditional systems of medicine (Ayurveda, Sidha and Unani). Turmeric was recommended as a specific single drug in Ayurveda for treating diabetes, leprosy, elephantiasis, extreme thirst and calculus. Since turmeric is endowed with multiple biological properties, it is being used to treat numerous ailments including common cold, allergy, skin diseases, inflammation, diabetes, high blood pressure, high cholesterol, jaundice, leprosy, malaria, ulcer, cardiac complaints and cancer. It is also used as an antidote for snake venom, as a veterinary medicine and as an insect and ant repellent. It is also used in various religious and magical rites (Ravindran et al., 2007).

India harbours rich diversity of turmeric and is cultivated as an annual crop for harvesting the rhizomes. It grows upto a height of 1.0 m and can be grown in diverse tropical conditions from sea level to 1500 m above sea level at a temperature range of 20-30ºC with an annual rainfall of 1500 mm or under irrigated conditions. Turmeric is raised in flat beds or in ridges either as a monocrop or intercropped with other plants under either rain-fed or irrigated condition (Babu et al., 1993). Traditionally, turmeric is propagated using rhizomes as seed set is rare in this crop due to self incompatibility. However, viable seed set was reported in turmeric and, IISR Prabha and IISR Prathibha were the first open-pollinated progenies of turmeric (Sasikumar et al., 1996). Although turmeric is cultivated in many countries, India is the largest producer and exporter of turmeric with major source from Andhra Pradesh, Tamil Nadu and Orissa. During 2015-2016, 943330 tonnes of

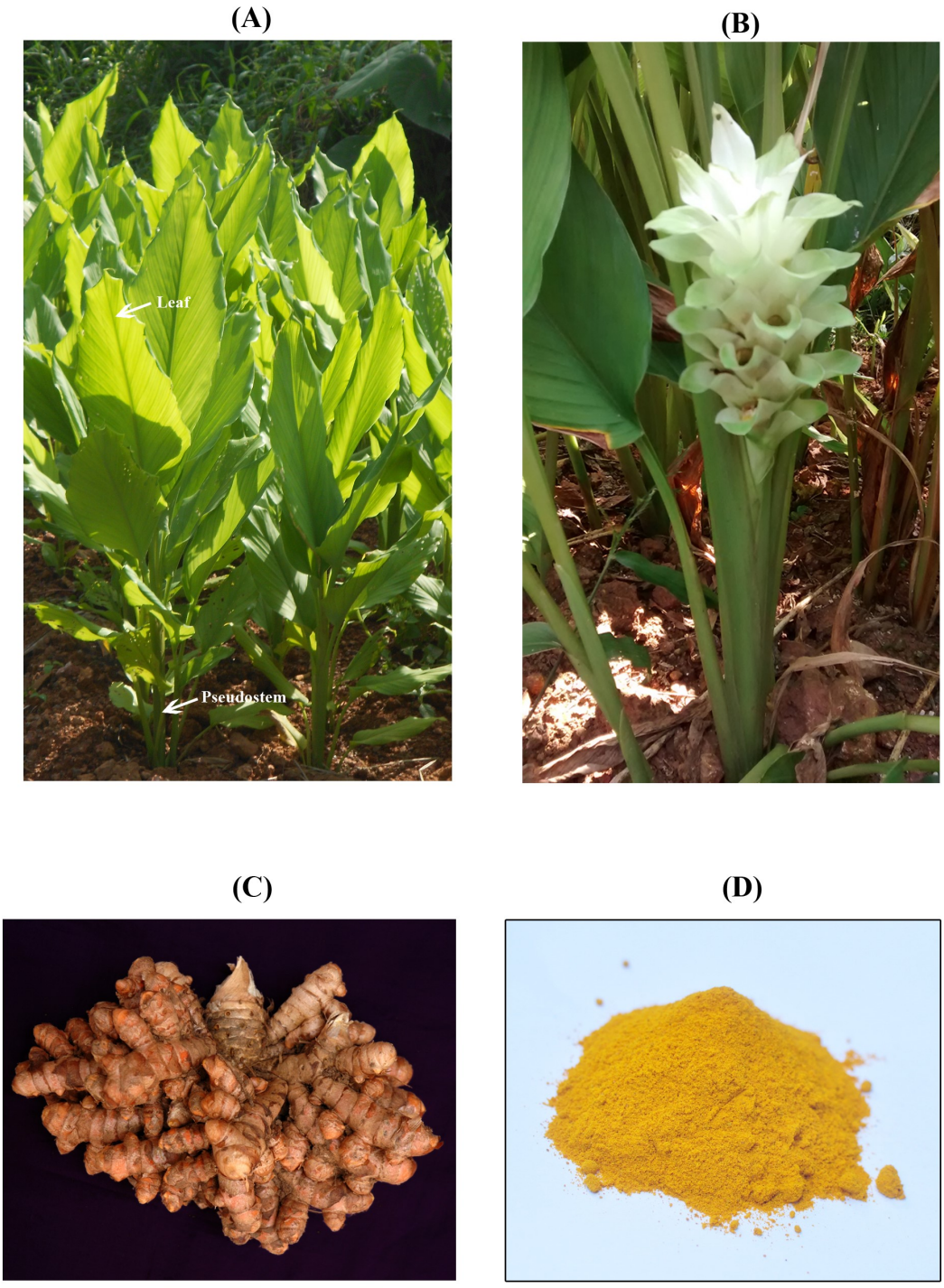


Figure 1. (A) Habit of *Curcuma longa* (turmeric), (B) turmeric flower, (C) turmeric rhizomes and (D) turmeric powder

turmeric was harvested from 185580 ha of which 88500 tonnes were exported from India (dasd.gov.in).

Turmeric is a triploid species (2n=3x=63) with a basic chromosome number, x=21. Inconsistency in chromosome number was also reported in turmeric with 2n=32, 2n=48, 2n=84, 2n=62 and 2n=64 (Sato, 1948; Raghavan and Venkatasubban, 1943; Sharma and Bhattacharya, 1959; Chakravorti, 1948; Nair and Sasikumar, 2009). Turmeric rhizomes are rich in carbohydrates, curcuminoids and essential oil. Turmeric oil is used in aromatherapy and perfume industry (Sasikumar, 2005; Babu et al., 1993). Curcuminoids and terpenoids are considered as the major bioactive compounds in turmeric (Lee et al., 2014) and the yellow colour of the rhizomes is attributed to curcuminoids (Chempakam and Parthasarathy, 2008). Curcuminoid mainly comprises curcumin (77%), demethoxycurcumin (17%) and bisdemethoxycurcumin (3%) (Goel et al., 2008; Strimpakos and Sharma, 2008) (**Fig. 2**).

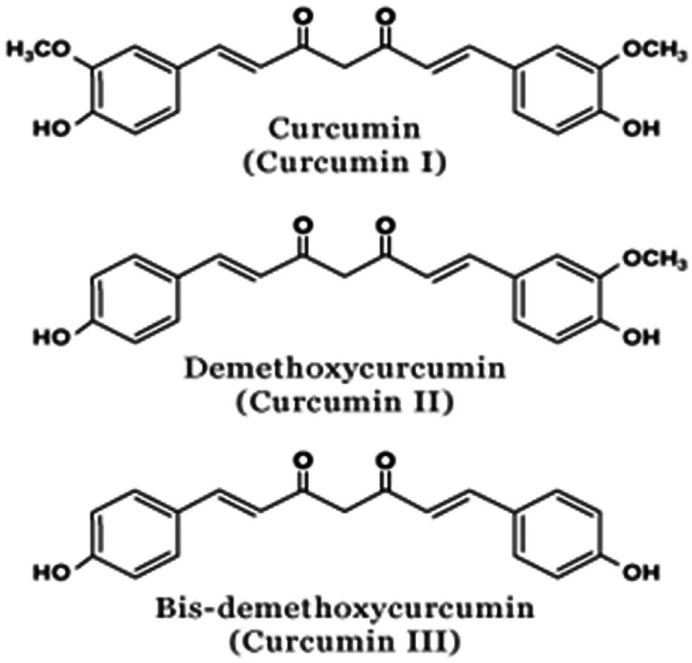


Figure 2. Chemical structure of curcuminoids (Chempakam and Parthasarathy, 2008)

2.2 Plant secondary metabolites

Plants produce an array of metabolites which are classified into primary and secondary metabolites. Primary metabolites are those vital for plant existence and secondary metabolites are those derived from primary metabolites and involved not only in defence mechanism against biotic and abiotic stresses but also to attract pollinators, to

adapt to new habitat and for successful reproduction (Bais et al., 2003; Dudareva et al., 2004; Hartmann, 2007). However, in many cases, the function of secondary metabolites and its benefit to the organism is not yet known (Raja and Sreenivasulu, 2015). Many of the biological activities of medicinal plants, which are used from centuries in traditional medicines are attributed to the plant secondary metabolites and are now being used as pharmaceutics, cosmetics and nutraceutics (Makkar et al., 2009; Bourgaud et al., 2001; Sadiq et al., 2014). Examples of plant derived drugs are artemisinin (Njuguna et al., 2012), vincristine (Eden et al., 2010), ginkgolide (Wang et al., 2006) etc. Based on their biosynthetic pathway, secondary metabolites were broadly classified into terpenoids (terpenes and steroids), alkaloids and phenylpropanoids (phenolics) (Zhao et al., 2013; Bourgaud et al., 2001). Phen ylpropanoids synthesized by fruits, vegetables, tea, cocoa etc. serve as a rich source of secondary metabolites which were reported to have diverse health benefiting properties like anti-inflammation, antioxidant, anticancer and antimicrobial activities and are involved in the biosynthesis of lignin and polyphenols like anthocyanin, tannin, resveratrol, coumarin, stilbene, flavonoids etc (Park et al., 2001; Vogt, 2010; Fraser and Chapple, 2011; Hassanpour et al., 2011). However, the low concentration of secondary metabolites and specificity in its synthesis (tissue or developmental stage specific) lead to its shortage and high price. This could be overcome by synthesizing it in elevated quantities through plant metabolic engineering after analyzing the complex pathway involved in secondary metabolite biosynthesis through extensive research (Li et al., 2014).

2.3 Biosynthesis and regulation of phenylpropanoids

The diversity in phenylpropanoids are attributed to specific enzymes organized in small gene families utilizing limited set of substrates derived from shikimate pathway. Phenylalanine ammonia lyase (PAL) catalyzes the deamination of phenylalanine, a product of shikimate pathway to cinnamic acid, the first step in phenylpropanoid biosynthesis. The hydroxylation of cinnamic acid by cinnamate 4-hydroxylase (C4H), a cytochrome P450-dependent monoxygenase yields *p*-coumarate. 4-coumarate: CoA ligase (4CL) catalyzes the conversion of *p*-coumarate to *p*-coumaroyl CoA, which is the direct precursor for synthesis of flavonoids, resveratrol, hydroxyl phenol lignin etc. The activity of hydroxycinnamoyl CoA transferase (HCT) on *p*-coumaroyl CoA yields *p*-coumaroyl shikimate which is acted upon by *p*-coumarate 3-hydroxylase (C3H) and hydroxycinnamoyl CoA transferase (HCT) to yield caffeoyl CoA, which is methylated by

caffeoyl CoA O-methyltransferase (COMT) to form feruloyl CoA. Feruloyl CoA is identified as the precursor for guaiacyl-lignin, syringyl-lignin and coumarin biosynthesis (Vogt, 2010; Baxter and Stewart Jr, 2013).

In dicots, PAL deaminates phenylalanine efficiently, whereas in monocots like maize, PAL deaminates both phenylalanine and tyrosine. PAL having tyrosine ammonia lyase (TAL) activity converts tyrosine into *p*-coumaric acid omitting the requirement of C4H (Rosler et al., 1997). Four isoforms of PAL was identified in *Arabidopsis*, five in poplar and nine in rice, each respond differentially to different stress conditions and were regulated at both developmental and spatial levels (Bhuiyan et al., 2009; Lillo et al., 2008). Most of the genes of secondary metabolite biosynthesis exist in families (Ober, 2005) or are arranged as clusters (Chu et al., 2011). Small gene families encoding enzymes of phenylpropanoids and lignin biosynthesis were identified in *Arabidopsis* and poplar (Hamberger et al., 2007).

Among the vast pool of secondary metabolites, phenylpropanoid derivative-flavonoid/anthocyanin biosynthesis is well characterized in many crops. Anthocyanins are pigments contributing to red, blue and purple colour of the plant tissues. Apart from its several biological functions such as tolerance to various stresses and attracting pollinators, they possess many human health benefiting properties (Tanaka et al., 2008). All the structural genes of anthocyanin biosynthesis were identified in many crops (Guo et al., 2014; Sun et al., 2016) and the sequences were found to be conserved across the plant kingdom (Holton and Cornish, 1995). Anthocyanin biosynthesis is divided into three phases; phenylpropanoid pathway, early steps of flavonoid pathway and late steps of anthocyanin specific pathway. In *Arabidopsis*, knockout mutant analysis and gene expression analysis have shown that two isoforms of *PAL* (*PAL1* and *PAL2*) (Olsen et al., 2008) and one isoform of *4CL* were involved in anthocyanin biosynthesis (Ehlting et al., 1999). More than one isoforms were reported for most of the anthocyanin biosynthetic pathway genes (Holton and Cornish, 1995; Jaakola et al., 2002). However, in *Arabidopsis*, DFR and ANS were encoded by single genes and in knockout mutants of *DFR* or *ANS*, anthocyanin was not detected (Shirley et al., 1995). The protein complex comprising enzymes of anthocyanin biosynthesis are associated with the membrane of endoplasmic reticulum for channelling the intermediates without diffusing it into cytoplasm (Jorgensen et al., 2005). The expression of anthocyanin biosynthetic pathway genes were found to be

regulated by a ternary complex composed of MYB, bHLH and WD40 transcription factors (Shi and Xie, 2014). The overexpression of MYB75, MYB90, MYB113 and MYB114 lead to high accumulation of anthocyanin and their knockout mutants lack anthocyanin. The gene expression analysis showed that the expression of DFR and ANS were higher in plants overexpressing these transcription factors and reduced in knockout mutants (Gonzalez et al., 2008). Three bHLH namely, GL3, EGL3 and TT8 were identified as the activators of anthocyanin biosynthesis whereas, only one WD40 protein namely, TTG1 was identified as regulator of anthocyanin biosynthesis in *Arabidopsis*. Two negative regulators, CPC and MYBL2 were also identified in *Arabidopsis* and the quantitative competition between positive and negative regulator determine the expression of structural genes. Two miRNAs, miR156 and miR828 were also identified as the regulators of anthocyanin biosynthesis (Shi and Xie, 2014). Similarly, in grapes, two MYBs (VvMYBA1 and VvMYBA2), one bHLH (VvMYCA1) and one WD32 (VvWDR1) regulate the anthocyanin biosynthesis (Xie et al., 2011).

2.4 Polyketide synthases and diarylheptanoid biosynthesis

Polyketides are a group of secondary metabolites and are synthesized using polyketide synthases (PKSs). PKSs are comprised of three groups; type I, type II and type III (Lussier et al., 2012). Plant PKSs are type III category and exist as homodimers with each monomer having active site catalysing the iterative decarboxylative condensation of malonyl CoA and starter CoAs generated from phenylpropanoid pathway (Dibyendu, 2015). Chalcone synthase was the first reported PKS from plants (Reimold et al., 1983). Thereafter, a number of type III PKSs were reported from plants which include those involved in biosynthesis of stilbenes, chromones, acridones, resorcinols, biphenyls, bibenzyls, benzophenones, pyrones etc. (Abe and Morita, 2010). They share 45-95% similarity and accept a wide range of substrates. The diversity of PKSs is on account of its specificity to substrates, number of condensation reaction and cyclization mechanism (Austin and Noel, 2003). Diarylheptanoids are phenylpropanoid derived metabolites (Vogt, 2010) and Schroder (1997) proposed that type III PKSs are involved in the biosynthesis of diarylheptanoids including curcuminoids.

Although they share similarity with each other, differences in a few amino acids resulted in different PKSs with distinct functions. Pentaketide chromone synthase (PCS) and octaketide synthase (OKS) of *Aloe arborescens* share high sequence similarity but

having different function. Both uses malonyl CoA as one of the substrates and PCS uses five malonyl CoA while OKS uses eight units, ie they differ in the number of condensation reactions. Thr 197, Gly 256 and Ser 338 are the active sites reported in chalcone synthase (CHS2) of *Medicago sativa*. In PCS, these active sites are replaced by Met, Leu and Val, while in OKS it is Gly, Leu and Thr. A single point mutation of Met corresponding to Thr 197 by Gly in PCS converts it to OKS. Similarly, the replacement of Gly corresponding to Thr 197 by Met in OKS converts it to PCS (Morita et al., 2007). Another example is 2-pyrone synthase (2-PS) of *Gerbera hybrida*, where the active sites are Leu, Leu and Ile corresponding to Thr 197, Gly 256 and Ser 338 of CHS2 in *M. sativa*. Replacing Thr 197, Gly 256 and Ser 338 of *M. sativa* with Leu, Leu and Ile in corresponding positions resulted in the conversion of CHS into 2-PS (Jez et al., 2000). A single mutation of Ser by Val in benzalacetone synthase of *Ruta graveolens* increased 2-fold activity of enzyme (Abe et al., 2007).

Apart from this, both chalcone synthase (CHS) and stibene synthase (STS) uses same substrates ie., three molecules of malonyl CoA and one molecule of *p*-coumaroyl CoA but differ in their cyclization mechanism to yield chalcone and resveratrol respectively. Cloning and sequence analysis of CHS and STS revealed no difference in their substrate binding residues, but differ in their residues lining the cyclization pocket (residues 262-268) and solvent exposed region (residues 253-259). Thus the product selectivity may be influenced by surface topology of the cyclization pocket which could alter the cyclization reaction (Ferrer et al., 1999).

2.5 Importance of curcumin

Curcumin was first identified by Vogel and Pelletier in 1815. The pure curcumin was extracted in 1842 (Vogel, 1842). Its chemical formula is $C_{21}H_{20}O_6$ and its chemical structure is diferuloylmethane or 1, 7-bis (4-hydroxy-3-methoxyphenyl) hepta-1, 6-diene-3, 5-dione (Milobedzka and Lampe, 1910). The structure of curcumin is similar to diarylheptanoid (Jayaprakasha et al., 2002) and is a non-flavonoid polyphenol (Tsao, 2010). Curcumin is insoluble in water but soluble in organic solvents and was approved by U.S. Food and Drug Administration as “Generally Recognized As Safe” (GRAS) (Prasad et al., 2014; Fadus et al., 2016). Dihydrocurcumin (DHC), Tetrahydrocurcumin (THC), hexahydrocurcumin (HHC), octahydrocurcumin (OHC), curcumin glucuronide, DHC-

glucuronide, THC glucuronide, and curcumin sulphate are the curcumin metabolites. Among these, THC was reported to be more stable than curcumin (Pan et al., 1999).

"Curcumin is a highly pleiotropic molecule with numerous targets and mechanisms of action" (Fadus et al., 2016) and the effectiveness of curcumin against cardiovascular diseases, inflammatory diseases, cancer, lung diseases, neurological diseases, liver diseases, metabolic diseases and autoimmune diseases was reviewed by Prasad et al. (2014). However, poor bioavailability of curcumin is a major constraint in developing it as a drug (Prasad et al., 2014; Fadus et al., 2016). The co-administration of piperine obtained from *Piper nigrum* and *Piper longum* increased the bioavailability of curcumin by 154% (Shoba et al., 1998). The immunoprotective effect of turmeric crude extract was reported to be more efficient than purified curcuminoid, which was suggested due to the synergistic activity of curcumin with other diverse consituents of turmeric extract (Abu-Rizq et al., 2015). A recent US patent application numbered U.S. Ser. No. 14/652,492 entitled "A novel composition of curcumin with enhanced bioavailability" discloses a composition containing curcumin mixture (curcumin dry crystals, volatile oil, fixed oil) and water extract (soluble proteins, dietary fibers, carbohydrates) in the ratio of 70:30 enhanced the bioavailability of curcumin. It also contained a natural emulsifier isolated from *Quillaja saponaria* and lecithin. The polysaccharide fraction of turmeric extract (NR-INF-02) was reported to be effective in acute and chronic inflammation (Illuri et al., 2015). A recent review by Nelson et al. (2017) reported that curcumin and its analogs do not meet the properties of a good drug candidate due to its less stability, reduced bioavailability and broad range of target activity. They suggested that the beneficial activity of curcumin may be due to its association with gut microbiome. The change in the gut microbiota with increased survival from cancer and Alzheimer's disease after the administration of curcumin was reported by McFadden et al. (2015) and Chin et al. (2013).

2.6 Genotype x environment interaction on curcumin content in rhizomes

The curcumin content was reported to vary within different accessions of *C. longa* (Wang et al., 1999; Jayaprakasha et al., 2002) and even when the same variety was planted in different agro-climatic conditions (Anandaraj et al., 2014). Shanmugasundaram et al. (2001) performed a trial with fifteen genotypes of turmeric at Tamil Nadu Agricultural University, Coimbatore. The curcumin content varied from 1.23 to 5.67% with highest curcumin content for RH-5 (5.67%). Boron deficiency was reported to increase the

curcumin content in rhizomes but it reduced the net yield and curcumin content of the plant (Dixit et al., 2002). HPLC based analysis of curcuminoids from four genotypes of turmeric collected from Salem, Erode, Mysore and Balasore showed that total curcuminoid content varied from 2.1 to 9.4%, with maximum reported for Salem genotype (9.4%). In all these samples, the percentage of curcumin was higher followed by demethoxycurcumin and bisdemethoxycurcumin (Jayaprakasha et al., 2002). A turmeric pot trial in Japan by Hossain and Ishimine (2005) reported that curcumin content was highest when the rhizomes were planted in dark red soil (pH- 5.2) than grey soil (pH-7.4) and red soil (pH-4.4). A field trial at ICAR Research Complex for North Eastern Hill Region, Umiam, Meghalaya with Megha Turmeric-1 variety cultivated under different management system showed that organic manure, specifically neem shield was better for enhancing curcumin content (6.76%) than 100% NPK (6.27%) and control (5.02%) (Sanwal et al., 2007).

The influence of sodicity on curcumin content was studied by Garg (2011) with four varieties of turmeric namely; Local BRS, Rajendra Sonia, KTS-1 and Meducar at Banthra Research Station of National Botanical Research Institute, Lucknow and reported that the curcumin content of Local BRS and Meducar variety increased with sodic soil condition while a reduced curcumin content was noted for Rajendra Sonia and KTS-1 in sodic soil. A mass field trial with eleven genotypes of turmeric in five environmental conditions across India by Anandaraj et al. (2014) reported that 17.7% of variation on curcumin content was attributed to environment and IISR Kedaram was found to be the most stable cultivar for curcumin production across these five environmental conditions. Low salinity (up to 60 mM) was found to increase the curcumin content; however a further increase in salinity resulted in drastic reduction in curcumin content (Mostajeran et al., 2014). Field trial with turmeric genotypes including Megha Turmeric and Roma at Mizoram identified Megha Turmeric as the good candidate for curcumin (Singh and Ramakrishna, 2014) whereas Roma was superior to Megha Turmeric when planted at Andhra Pradesh (Kumar et al., 2015). A study conducted with turmeric variety, Roma in nine agro-climatic conditions of Odisha showed that increase in organic carbon from 0.48-0.965% and potassium from 309-388.4 kg/h increased the curcumin content from 3.2-3.8% and 2-5% respectively. However, excess amount of organic carbon (1%) and potassium (743.7 kg/h) reduced the curcumin content to 1.7% and 3.2% respectively (Sandeep et al., 2015)

Indian turmeric has great value in market due to its high curcumin content. All India Coordinated Research Project on Spices (AICRPS), an Indian Council of Agricultural research (ICAR) project is the largest spices research network in the country with its headquarters at ICAR - Indian Institute of Spices Research (IISR), Kozhikode, Kerala. A National conservatory of turmeric germplasm with 1404 accessions is maintained at ICAR-IISR, Kozhikode, Kerala. Based on field evaluation for yield parameters and curcumin content (%), eight varieties of turmeric namely, IISR Pragathi (5%), IISR Prathibha (6.52%), IISR Prabha (6.52%), Suvarna (4.0%), Suguna (4.9%), Sudarshana (7.9%), Kedaram (5.7%) and IISR Alleppey Supreme (5.5%) have been released from IISR. The recently released IISR Pragathi is a short duration variety with high yield and satble curcumin content across locations (www.spices.res.in). Screening for the presence of secondary metabolites in the ethanolic extract of turmeric identified the presence of phenols, flavonoids, alkaloids, terpenoids, tannins and saponins (Dutta, 2015).

2.7 Techniques applied to decipher the biosynthetic pathway of secondary metabolites

2.7.1 Precursor feeding studies

Precursor feeding is one of the techniques to over express the genes involved in biosynthetic pathway (Sivanandhan et al., 2014). The addition of precursors at optimum concentration can stimulate the production of desired secondary metabolite in *in vitro* conditions (Bourgaud et al., 2001). Here the precursors are fed into the tissue culture plant media and the metabolites were compared with the control. The addition of phenylalanine into the hairy root cultures of *Silybum marianum* resulted in enhanced accumulation of silymarin (Rahimi et al., 2011). Similarly, coniferyl alcohol increased the production of podophyllotoxin (Exposito et al., 2009). Radioactive/stable isotope-labelled precursor study was utilized to elucidate taxol, artemisinin and anthocyanin biosynthesis (Li et al., 2014). When ^{13}C-labelled phenylalanine was added to the *Vitis vinifera* cell suspension cultures, more than 50% of labelled phenylalanine was incorporated into anthocyanin (Krisa et al., 1999).

2.7.2 Metabolomics approach

Metabolomics is the comprehensive analysis of all or many metabolites in a biological sample at a particular time point under specific conditions. Metabolite analysis

mainly uses mass spectroscopy based platforms. Silencing of target gene followed by comparison of its metabolic profile with that of wild type genotype identified many novel genes and routes of biosynthetic pathway (Nguyen et al., 2012). The comparison of metabolites from wild type with mutant, stressed condition with control and different tissues of same plant identified differential and many novel metabolites in plants. For example, among the four isoforms of PAL in *Arabidopsis thaliana*, *PAL1* has been identified as involved in phenylpropanoid biosynthesis through gene knock out and metabolome studies (Rohde et al., 2004). Comparative metabolomic profile of resistant and susceptible genotypes of wheat to *Fusarium graminearum* identified many phenylpropanoid metabolites that were either induced or up-regulated in resistant phenotype (Gunnaiah et al., 2012). Tissue specific metabolome analysis of *A. thaliana* identified flavonol glycoside in flowers as a tissue specific metabolite (Matsuda et al., 2010). Metabolomic analysis of soybean leaves during phase transition from vegetative to reproductive stage identified eight flavonoid kaempferol glycosides as growth markers (Song et al., 2014). However, limited number of compound identification (80-200 compounds) by mass spectroscopy based platforms and difficulty in analysing unknown chromatographic peaks is challenging in metabolomic studies (Zhao et al., 2013).

2.7.3 Transcriptome analysis

Transcriptome is the complete set of transcripts in a cell, including mRNAs, noncoding RNAs and small RNAs (Wang et al., 2016). In many non model plants, homology based gene cloning and transcriptome library analysis paved the way to identify the genes of secondary metabolite biosynthetic pathway. In homology based gene mining, conserved regions of the target gene sequences available from closely related species are used to design degenerate primers, which are used to isolate the desired gene using PCR strategy or hybridization method from the crop of interest (Zuiter et al., 2012; Bogs et al., 2005). This approach was found to be successful in many crops where the target biosynthetic pathway genes were already well elucidated in model crops. For example, Wei et al. (2011) amplified six anthocyanin biosynthetic genes from *Litchi chinensis* using degenerate primers and studied their expression pattern using real time PCR across 12 genotypes of varying anthocyanin content. Correlation analysis was performed between gene expression and anthocyanin concentration. The results showed that the expression of UDP-glucose: flavonoid 3-O-glucosyltransferase (UFGT) correlated significantly with

anthocyanin content. They suggested that UFGT played a dominant role in anthocyanin accumulation in litchi. Shafrin et al. (2017) used degenerate primers to amplify cinnamate 4-hydroxylase (*C4H*) and caffeic acid O-methyltransferase (*COMT*) from jute (*Corchorus olitorius*). To reduce lignin content in jute, *C4H* and *COMT* sequences were used to construct hairpin RNA and transgenic plants expressed reduced level of lignin. Karamat et al. (2014) applied degenerate primers to amplify prenyltransferase gene involved in furanocoumarin, a subgroup of coumarin biosynthesis from parsley. Full length cDNA of prenyltransferase was isolated using 5' and 3' RACE (Random amplification of cDNA ends) and analyzed its expression using real time PCR in different tissues. In plants treated with UV-B, there was a significant increase in the expression of prenyltransferase gene and furanocoumarin content. Transcriptome library based studies need not require the prior knowledge of the plant's genome. In this study, the cDNAs were prepared and sequenced either after cloning into a suitable host (Expressed Sequence Tag) or using next generation sequencing platforms (Zhao et al., 2013). The abundance of transcript could also be evaluated from the transcriptome. Comparing the transcriptomes, developed from samples with differential metabolite contents (curcumin), putative genes involved in biosynthesis could be identified. Further, the data could be validated using DNA blot or northern blot or virtual northern blot or real time PCR assays (Li et al., 2016; Coelho et al., 2013; Schmidtke et al., 2011; Zhou et al., 2017).

Fridman et al. (2005) compared phage vector based transcriptomes constructed from two tomatoes differing in methylketone content and identified methylketone synthase correlating with methylketone. The recombinant methylketone synthase and its antibodies were synthesized. Ten different accessions of tomato were used to validate the results. Quantification of methylketone content by GC-MS, transcript abundance using northern blot and protein abundance using antibodies was performed. The results suggested methylketone synthase is a candidate gene involved in methylketone biosynthesis. Roslan et al. (2012) constructed three cDNA libraries from *Polygonum minus*; one standard library from leaves and two normalized full length library from root and stem respectively using lambda phage vectors. Sequencing of libraries identified 4196 expressed sequences tags (ESTs) in which 11 ESTs were mapped to seven genes of flavonoid biosynthetic pathway. Among them, the expression of chalcone synthase, flavonol synthase and leucoanthocyanidin dioxygenase were studied in leaves, stems and roots and it was observed that the expression was higher in roots as analyzed using real time PCR.

To study catechin biosynthesis, Park et al. (2004) isolated mRNA from young and mature leaves of *Camellia sinensis* (tea) and catechin was reported to be higher in young leaves. A subtractive cDNA library was constructed from these samples and ligated to pGEM-T Easy vector. Sequencing of 558 clones revealed the presence of flavonoid biosynthetic genes namely, chalcone synthase, flavanone 3-hydroxylase, flavonoid 3,5-hydroxylase, flavonol synthase, dihydroflavonol 4-reductase and leucoanthocyanidin reductase. Quantitative RT-PCR and DNA blot assay showed that the expression of flavanone 3-hydroxylase, dihydroflavonol 4-reductase and leucoanthocyanidin reductase were higher in young leaves. Similarly, Park et al. (2007) compared the transcripts of green and red lettuce leaves using subtractive cDNA approach. Dot-blot analysis was done for 566 clones and found that 53 transcripts were over-expressed in red leaves. Northern blot analysis was done for chalcone synthase, flavanone 3-hydroxylase, flavonoid 3,5-hydroxylase, dihydroflavonol 4-reductase and anthocyanidin synthase. The expression of three genes namely, chalcone synthase, flavanone 3-hydroxylase and dihydroflavonol 4-reductase showed positive correlation with anthocyanin accumulation in UV-B irradiated lettuce leaves.

The transcripts associated with drought stress were identified in *Lablab purpureus* using suppression subtraction hybridization approach by Yao et al. (2013) where ESTs were generated from drought tolerant variety subjected to water stress after 10 days of germination and compared with that of well watered control for 10 days. Reverse northern blot was done to screen the transcripts with 1.5 fold expression and 1525 clones with ≥ 1.5 fold expression were sequenced. Real time PCR with 10 transcripts was carried out to validate suppression subtraction hybridization data and the results showed that these transcripts were up-regulated during water stress. Guo et al. (2013) studied the role of abscisic acid in alleviating the chilling related injury in *Capsicum annuum* using suppression subtraction hybridization method. Of the 18 selected transcripts, 10 were found to be up-regulated and eight were down-regulated in abscisic acid pre-treated seedlings under chilling stress when compared with control seedlings and thus suggested that abscisic acid could regulate the expression of genes under chilling stress.

2.7.3.3 Application of high throughput sequencing in transcriptome analysis

The completion of Human Genome Project revealed the necessity of advanced technologies to resolve the complexity of biological samples. However, high cost of

sequencing and limited throughput remained as the major barriers. With the advent of next generation sequencing platforms like Roche 454, SOLiD, Illumina etc., the cost of sequencing has reduced substantially and millions of short DNA fragments could be sequenced in short time. High throughput (formerly 'next generation') sequencing platforms provide vast amount of data, but due to generation of contigs through alignment of shorter reads of 35-700 bp, the associated error rates (~0.1-15%) are higher than that of traditional Sanger sequencing platform and hence the data analysis remained as a bottleneck in understanding the genomics (Goodwin et al., 2016).

The two approaches for the generation of short reads in high throughput sequencing are sequencing by ligation (SBL) and sequencing by synthesis (SBS). In SBL approach, a fluorophore bound probe hybridizes with a DNA fragment and ligates to adjacent oligonucleotide for imaging. The identification of nucleotide base is based on the emission spectrum of fluorophore. SBS approach utilizes a polymerase and the incorporation of nucleotide into an elongating strand is identified based on the signal such as a fluorophore or change in ionic concentration. SBS approach will be classified into cyclic reversible termination and single-nucleotide addition. In both approaches, DNA is fragmented followed by the addition of specific adaptors to both 5' and 3' ends and its clonal amplification on a solid surface. A sequencing platform can collect information from millions of SBL or SBS reaction centres resulting in massive parallel sequencing.

Roche 454 was the first commercially successful next generation system and is based on sequencing by synthesis-single nucleotide addition principle. Here, the template bound beads are distributed into a PicoTiterPlate along with the primer and enzyme cocktail containing DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5′ phosphosulfate and luciferin. DNA polymerase elongates the complementary strand if the added dNTP pairs with the template and releases pyrophosphate. ATP sulfurylase converts adenosine 5′ phosphosulfate to ATP which in turn is a cofactor, converting luciferin to oxyluciferin by luciferase, resulting in a bioluminescence signal. The signal is detected by charge-coupled device camera, attributing to the incorporation of base. Meanwhile, the unmatched and excess dNTPs are degraded by apyrase. Another dNTP is added to the system and the reaction is repeated (Fakruddin et al., 2012). Sequencing by Oligo Ligation Detection (SOliD) system employs sequencing by ligation principle (SBL) in which a probe of eight nucleotide bases having two fluorophore-labelled nucleotides followed by

six degenerate bases which is ligated to an anchor, complements with the adaptor. Four fluorescent signals each representing a subset of four dinucleotide combinations are used and hence each ligation signal represents one of the possible dinucleotide. The first two bases in each fragment were then identified followed by cleavage at 5th base of the probe. The process is repeated four times until two out of every five bases are identified (Goodwin et al., 2016). Illumina platform adopts the technology of sequencing by synthesis-cyclic reversible termination. The denatured DNA with adaptors is loaded to the flow cell and clonal DNA fragments are synthesized based on bridge amplification. Prior to sequencing, the library splices into single strand with the help of linearization enzyme and during each cycle, mixture of four individually labelled and 3´ blocked dNTPs are added which could complement the template one base at a time followed by detection of signal by charge-coupled device camera (Liu et al., 2012).

Pathak et al. (2013) utilized 454 GS FLX platform to develop two transcriptomes from *Papaver somniferum* with contrasting papaverine content to elucidate papaverine biosynthetic pathway genes. The transcriptomes were compared and they suggested that (S)-coclaurine is an intermediate for papaverine biosynthesis. Gupta et al. (2013) also utilized 454 pyrosequencing to deduce the genes involved in withanolide biosynthesis in *Withania somnifera*. Transcriptomes generated from leaves and roots which synthesize withaferin A and withanolide respectively were studied and it was suggested that cytochrome P450, glycosyltransferase and methyltransferase are involved in tissue specific expression of withanolides. Further Gupta et al. (2015) used different chemotypes of *Withania somnifera* and reported the differential expression of cytochrome P450, glycosyltransferase, methyltransferase and transcription factors that might be involved in chemodiversity in *W. somnifera*. Gordo et al. (2012) developed root specific transcriptome of black pepper using SOLiD platform and identified 615 root specific proteins. Zhang et al. (2015) utilized SOLiD system of sequencing to investigate the molecular mechanism underlying flower development in *Arabidopsis* and identified 23961 florally expressed genes and many floral development specific genes. Wei et al. (2015) studied the differential expression of anthocyanin biosynthetic genes in four stages of fruit ripening in red and yellow fruits of *Prunus avium* using Illumina platform. Eighteen structural and regulatory unigenes were identified and further validated using quantitative real time PCR. Torre et al. (2016) utilized Illumina platform to develop transcriptomes from leaves of green and red basil (*Ocimum basilicum*) after exposing the plants to sunlight for four

weeks and found that biosynthesis of waxes was predominant in red basil while in green basil; transcripts involved in flavonol and carotenoid biosynthesis were up-regulated.

Among the three commercial sequencing platforms, Illumina has high throughput and cost effectiveness whereas, SOliD has highest accuracy and Roche 454 system has longest read length (Goodwin et al., 2016). Xiao et al. (2013) compared Roche 454 and Illumina platform of sequencing in 20 plant species and reported that Illumina platform produced 100-fold more reads than Roche 454 pyrosequencing and hence more contigs were produced from Illumina platform. The identification of full length unigenes were higher for assemblies generated from Illumina sequences (63.2%) compared with 454 dataset (48.1%). However, mining of full length sequences was 40% higher when these databases generated from Illumina and Roche 454 platform were combined. Gordo et al. (2012) reported that the number of contigs were low from SOLiD sequencing platform and could annotate only 50-60% of reads against a transcriptome reference. Weber (2015) reported that 454 pyrosequencing has been mainly superseded by Illumina sequencing due to its enormous throughput and the dominant technology in the sequencing market.

2.7.4 Bioinformatic approaches

Molecular docking is a convenient method to study protein-ligand interaction (Ogungbe and Setzer, 2016) where the protein-ligand orientation was envisaged to predict the affinity of ligand and hence used to rank the ligands. Homology modelling was applied to the proteins with high sequence homology to known 3D structural data generated through X-ray crystallography and NMR spectroscopy (Monika et al., 2010). Molecular docking was applied to channelize the precursor 2,3-oxidosqualene to triterpenoid biosynthesis in *Centella asiatica*. Cycloartenol synthase and β-amyrin synthase are two isoforms of 2,3- oxidosqualene synthase, where the activity of cycloartenol synthase on 2,3-oxidosqualene leads to phytosterol biosynthesis and that of β-amyrin synthase on the same precursor 2,3- oxidosqualene channelize it to triterpenoid biosynthesis. Homology based modelling was applied to generate the 3D structures of cycloartenol synthase and β-amyrin synthase and were docked with a panel of ligands. The study showed that the ligand, ketoconazole binds effectively with cycloartenol synthase and not with β-amyrin synthase. The addition of ketoconazole in the cell cultures of *Centella asiatica* channelize the precursor 2,3- oxidosqualene to triterpenoid biosynthesis. Amino acid substitution

studies identified LYS 728 as the key amino acid for ketoconazole-cycloartenol synthase interaction (Kumar et al., 2013).

In another study by Awasthi et al. (2016) in *Coleus forskohlii*, two differentially expressed cytochrome P450 monooxygenases (CYP) exhibiting maximum expression in leaves compared to stem and root tissues were identified using degenerate primers. Full length cDNA was amplified using RACE kit and homology modelling was performed. Real time PCR studies showed that mannitol treatment increased the expression of CYP93B, while that of CYP706C was down-regulated. Metabolite analysis revealed that the contents of secondary metabolites, genkwannin was increased and that of anthocyanin was decreased. Molecular docking studies reported that CYP93B had higher affinity for genkwanin and CYP706C binds effectively with leucopelargonidin, an intermediate in anthocyanin biosynthesis. They suggested that mannitol treatment increased the genkwanin content by inhibiting competitive branch of flavonoids.

Bhat et al. (2013) identified two uridine diphosphate glycosyltransferases (UGTs) from *Picrorhiza kurrooa*. The full length cDNA was amplified using RACE kit and 3D structures were generated through homology modelling. Both UGTs showed differential expression in leaves, rhizomes and inflorescence. The upstream regions of both genes were amplified using universal genome walker kit. The presence of *cis*-regulatory elements correlated with differential gene expression pattern. Molecular docking studies suggested UGT94F2 as an iridoid-specific glucosyltransferase, while UGT86C4 was predicted to be a kaempferol-specific glucosyltransferase.

Salvinorin A, the active metabolite of *Salvia divinorum* was reported to be potential candidate in treating neuropsychiatric diseases and drug addictions. Leaf and trichome specific transcriptome of *S. divinorum* identified five diterpene synthases and among them, the putative role of clerodienyl diphosphate synthase in salvinorin A biosynthesis was suggested based on its molecular docking with its substrate geranylgeranyl diphosphate and further functionally characterized using tissue specific expression analysis, transient expression in *Nicotiana benthamiana, in vivo* co-expression and *in vitro* assays (Pelot et al., 2017).

In plants, tyrosine can be synthesized either using prephenate dehydrogenase or arogenate dehydrogenase. Using molecular docking studies, Schenck et al. (2017) reported

that alternative tyrosine biosynthetic pathway is attributed to the single amino acid residue (Asn 222) that defines substrate specificity and regulation. Mutant studies showed that all arogenate dehydrogenases had conserved aspartate at the corresponding position 222 whereas, replacement by other residues confer prephenate dehydrogenase activity.

2.7.5 Plant genetic manipulation

Over expression/silencing of gene are the two approaches to functionally characterize a particular gene. A fivefold increase in resveratrol content was observed by over expressing stilbene synthase gene in *Vitis vinifera* (Fan et al., 2008). Similarly, over expressing squalene synthase in *Panax ginseng* increased phytosterol and ginsenoside content (Lee et al., 2004). Apart from structural genes, transcription factors were also functionally characterized by this method. Over expression of MYB 113 or MYB 114 increased the anthocyanin accumulation in *Arabidopsis* (Gonzalez et al., 2008). Morphine biosynthesis was reduced by silencing codeinone reductase gene in *Papaver somniferum* using RNA interference technology (Allen et al., 2004). RNA interference technology was used to silence dihydroflavonol 4- reductase in strawberry, which resulted in reduced anthocyanin concentration (Lin et al., 2013). The accumulation of specific secondary metabolite could be enhanced by silencing the committed gene of undesired metabolite, thereby channelling the precursors to the specific secondary metabolite pathway. *p*-coumaroyl CoA formed from phenylalanine is the precursor for both anthocyanin and lignin biosynthesis. If *p*-coumaroyl CoA is acted upon by chalcone synthase (CHS), metabolic flux will be directed towards the anthocyanin production and if catalyzed by hydroxycinnamoyl transferase (HCT), it is directed towards lignin biosynthesis. The silencing of HCT resulted in plants with reduced lignin content and increased anthocyanin accumulation (Besseau et al., 2007). Similarly, the silencing of CHS produced transgenic plants with increased lignin content and reduced anthocyanin content (Eloy et al., 2016).

2.7.6 Proteomics

"Proteomics" coined by Wilkins et al. (1996) is the study of entire proteins expressed by a genome in a given time under defined conditions. When compared to genome, which is stable, proteome is complex as it changes with time and environmental conditions. Both gel based and gel free methods were employed to study and compare the proteomes. Proteomics has been applied to decipher genes and enzymes, mainly

transcription factors of secondary metabolic pathways (Aryal et al., 2014). Both gel-based and gel free approaches applied in proteomics was reviewed by Abdallah et al. (2012). Among them, two dimensional electrophoresis (2DE) has better resolution. In 2DE, the proteins are first focussed according to isoelectric point and then resolved in SDS-PAGE (PolyAcrylamide Gel Electrophoresis) based on its molecular weight.

The major hurdle in this approach is the lower abundance of proteins related to secondary metabolism and higher background noise of non-target proteins (Martinez-Esteso et al., 2015). The identification and quantification of relative abundance of phenylpropanoid and flavonoid pathway enzymes was carried out in grapes through berry skin proteome analysis (Martinez-Esteso et al., 2011). Similarly, proteomics of trichomes of tobacco identified enzymes involved in the synthesis of terpenoid precursors and transporters related to secondary metabolite transport (Van Cutsem et al., 2011). The limited amount of genomic and proteomic data in non model crops is a barrier to annotate the proteins of specialized pathway. For example, 19 differentially expressed proteins were identified by comparing the proteome of two developmental stages of *Euphorbia kansui*. However, only four of them had functional description (Zhao et al., 2014). In order to overcome this impediment, differential proteins were identified, digested with trypsin and subjected to LC-MS/MS to generate peptide sequence data. The peptide sequence was then used to query the specific EST database using Mascot tool for annotation. Two dimensional approach could lead to identification of only one enzyme of benzylisoquinoline alkaloid biosynthetic pathway in opium poppy (Zulak et al., 2009) while the integrated transcriptome and proteome approach identified all the enzymes in this pathway (Desgagne-Penixetal et al., 2010).

Isobaric tags for relative and absolute quantitation (iTRAQ) based gel-free proteomics approach is a recently developed technique which allows the unbiased detection of protein abundances within a cell and is suitable for identifying differentially expressed proteins under various developmental processes (Kambiranda et al., 2014) and in stress induced studies (Dong et al., 2013). Chen et al. (2016) utilized iTRAQ to compare the protein profile of five developmental stages of black rice (*Oryza sativa*) which is rich in anthocyanin and identified chalcone synthase as the key enzyme in the anthocyanin biosynthesis.

2.7.7 Integrated 'omics' approach to elucidate biosynthetic pathway

Integrating transcriptomics and metabolomics approach in *Arabidopsis* by Hirai et al. (2005) found that downstream genes of anthocyanin biosynthesis clustered with a regulatory gene *MYB75*, an activator type transcription factor. Tohge et al. (2005) identified two novel glycosyltransferase namely, flavonoid 3-O-glucosyltransferase and anthocyanin 5-O-glucosyltransferase with putative roles in anthocyanin biosynthesis using combined transcriptomic and metabolomic approach. Ge et al. (2015) identified 19 secondary metabolites through metabolite profiling by LC/MS and 25 differentially expressed transcripts by transcriptome sequencing in response to jasmonate induction in *Salvia miltiorrhiza*. This combined approach identified a repressor in the jasmonate signaling pathway. Zhao et al. (2015) identified 7 cytochrome-P450s, 8 multidrug resistance transporter unigenes and 8 biomarker compounds in the biosynthesis and transport of coumarin compounds in *Peucedanum praeruptorum* using Illumina based transcriptome database and HPLC coupled with electrospray-ionization quadrupole time-of-flight mass spectrometry (Q-TOF MS) based metabolomics dataset. Wan et al. (2017) could detect two promising genes underlying fatty acid biosynthesis pathway in *Brassica napus* as acetyl-CoA carboxylase and acyl-ACP desaturase from both Illumina based transcriptome sequencing and iTRAQ based proteomic approach and further confirmed through temporal gene expression analysis using real time PCR. Zhan et al. (2016) identified all the possible genes/proteins involved in triterpenoid saponin biosynthesis in *Anemone flaccida* using a combination of transcriptome and proteome based on Illumina HiSeq 2000 platform and iTRAQ technique respectively.

2.8 Biosynthesis of curcuminoids

Owing to the multiple pharmacological activity of curcumin, the urge to elucidate its biosynthetic pathway has been initiated many decades ago. The radiotracer studies by Roughley and Whiting (1973) reported that curcumin is a phenylpropanoid derivative as (1- and 3- ^{14}C) phenylalanine was detected in curcumin. Radiotracer studies were carried out at our institute by Neema (2005), who also reported that curcumin is a phenylpropanoid derivative as the incorporation of ^{14}C-phenylalanine label was detected in phenolic acids and even in curcuminoids. The study suggested that most of the ferulic acid is channelized to curcumin biosynthesis and phenylalanine ammonia lyase (PAL) is the key regulatory enzyme in curcumin biosynthesis. The study showed that enzyme activity

of PAL isolated from leaves significantly correlated with curcuminoid content in five released turmeric varieties. Ramirez-Ahmuda et al. (2006) provided the enzymatic evidence for the presence of phenylalanine ammonia lyase, hydroxycinnamoyl transferase, O-methyltransferase and curcuminoid synthase activity in the protein crude extracts of leaf, shoot and rhizome tissues. They suggested that either single or multiple polyketide synthases are involved in curcumin biosynthesis. The presence of hydroxycinnamoyl-CoA thioesterase was also reported from the crude extracts and suggested that thioesterases could have a regulatory role in curcuminoid biosynthesis by competing for the substrates (*p*-coumaroyl CoA and feruloyl CoA) thus preventing curcumin formation. Katsuyama et al. (2007) identified curcuminoid synthase from *Oryza sativa* catalyzing the formation of bisdemethoxycurcumin utilizing two molecules of *p*-coumaroyl-CoAs and one molecule of malonyl-CoA with diketide CoA intermediate. A recombinant *E. coli* harbouring phenylalanine ammonia lyase (PAL) from the yeast *Rhodotorula rubra*, 4-coumarate: CoA ligase (4CL) from *Lithospermum erythrorhizon* and curcuminoid synthase (CUS) from *Oryza sativa* was constructed and utilized for the production of bisdemethoxycurcumin with tyrosine or phenylalanine, or both as substrates (Katsuyama et al., 2008). ^{13}C labelled precursor study by Kita et al. (2008) showed that malonic acid or acetic acid and phenylalanine or phenylpropanoids were incorporated into demethoxycurcumin, whereas tyrosine was not integrated into curcuminoid skeleton. The precursor, cinnamoyl CoA had higher incorporation efficiency into demethoxycurcumin than *p*-coumaric acid and ferulic acid. They assumed that two molecules of cinnamoyl CoA and one molecule of malonyl CoA were utilized for curcuminoid biosynthesis and methoxy- and hydroxyl- groups were incorporated into the aromatic ring after the formation of curcuminoid skeleton.

However, metabolic profiling analysis by Xie et al. (2009) identified many co-regulated metabolite modules in turmeric in which one module had curcumin, demethoxycurcumin and bisdemethoxycurcumin and was distinct from other diarylheptanoids, which lead to a hypothesis that 3-methoxy group was added to the aromatic ring before the formation of curcuminoid skeleton. They also suggested that multiple PKSs with different substrate selectivities were involved in the formation of the twelve diarylheptanoids. One PKS was assumed to utilize caffeoyl CoA for the synthesis of 3'-hydroxy-bisdemethoxycurcumin, 3'-hydroxydemethoxycurcumin and 3'-hydroxy-6,7-dihydro-bisdemethoxycurcumin, whereas the second PKS used 5-hydroxyferuloyl-CoA as substrate for the production of 5'-hydroxycurcumin and 5'-hydroxy-

demethoxycurcumin. A third PKS catalyzed the formation of curcumin, demethoxycurcumin and bisdemethoxycurcumin but could not make use of either caffeoyl CoA or 5-hydroxyferuloyl-CoA as substrates.

Katsuyama et al. (2009a) designed primers from expressed sequence tag (EST) sequences of turmeric for the amplification of two PKSs; diketide CoA synthase (DCS) catalyzing the formation of diketide CoA utilizing *p*-coumaroyl-CoA, feruloyl-CoA and malonyl-CoA as substrates and curcumin synthase 1 (CURS1) catalyzing the formation of curcuminoids from diketide CoA by condensing it with the starter substrates. Full length cDNA of these two PKSs was amplified using RACE (Rapid amplification of cDNA ends) kit, ligated in pET16b expression vector and his-tagged recombinant protein was expressed in *E. coli* BL21 (DE3). In enzyme kinetic assay of DCS with cinnamoyl-CoA, *p*-coumaroyl-CoA, or feruloyl-CoA as substrates and malonyl-CoA as extender substrate, the presence of diketide-CoA was detected in the feruloyl CoA primed reaction. No diketide CoA was detected either in cinnamoyl-CoA or *p*-coumaroyl-CoA primed reaction. Trace amount of curcumin was detected when recombinant CURS1 was incubated with either feruloyl CoA or *p*-coumaroyl-CoA as starter substrates and malonyl CoA as extender substrate. Hundred fold curcumin was detected when CURS1 was incubated with cinnamoyl-diketide-N-acetylcysteamine and feruloyl CoA. The co-incubation of DCS and CURS1 with feruloyl CoA and malonyl CoA yielded curcumin and the addition of *p*-coumaroyl CoA to the same reaction could detect the presence of demethoxycurcumin and bisdemethoxycurcumin in trace amount. The relative expression analysis of *DCS* and *CURS1* by real time PCR analysis using four month old turmeric seedling showed that both had higher expression in rhizomes than leaves.

Two more PKSs ie., CURS2 and CURS3 were also identified in turmeric by following the same strategy followed for the identification of DCS and CURS1. No curcuminoids were detected when purified recombinant enzymes; CURS2 and CURS3 were incubated with feruloyl CoA and malonyl CoA. However, curcumin was detected when DCS was added to the reaction. The presence of bisdemethoxycurcumin was also detected when *p*-coumaroyl CoA was added to the same reaction. The substrate specificity analysis showed that CURS2 preferred feruloyl CoA as that of DCS and CURS1, while CURS3 preferred both feruloyl CoA and *p*-coumaroyl CoA. The relative expression analysis of *CURS2* using real time PCR reported its higher expression in rhizomes than

leaves as that of *CURS1* but *CURS3* had similar expression in both leaves and rhizomes of four month old turmeric seedlings (Katsuyama et al., 2009b).

Resmi and Soniya (2012) identified partial sequences of ten *PKS*s ie., *ClPKS1-10* by reverse transcription PCR using the primers designed from the conserved sequences of *PKS*s and homology based data mining from ESTs. Phylogenetic analysis of deduced amino acid sequences from partial cDNA sequences clustered ten clones into two groups i.e. chalcone and non chalcone forming ones. Among them full length cDNA of *ClPKS9* (similar to chalcone synthase) and *ClPKS10* (similar to curcumin synthase) were amplified. The relative gene expression analysis by real time PCR showed higher expression of *ClPKS9* in shoot and *ClPKS10* in leaves. They suggested that curcuminoids are synthesized not only in rhizomes but also in aerial parts.

Koo et al. (2013) utilized cDNA library synthesized from leaves, rhizomes and roots of turmeric and observed that clones harbouring phenylpropanoid biosynthetic pathway genes were abundant in all the tissues, with cinnamate 4-hydroxylase as the abundant gene. More than 40 unigenes were classified as PKSs which included those showing identity to chalcone synthase, diketide CoA synthase, curcumin synthases and other putative PKSs. The unigenes showing identity to diketide CoA synthase and curcumin synthase were abundant in all the tissues and they also suggested that large array of diarylheptanoids in turmeric is attributed to several PKSs which can catalyze different reactions utilizing different substrates. They also predicted that β-ketoacyl-CoA synthase-like subclass, specific reductases and hydroxylases may also have role in curcumin biosynthesis. The unigenes showing similarity to MYB, NAC, WRKY, homeobox, bZIP and CONSTANS were the major transcription factors identified from the library. MYB was the abundant class and presumed to have a role in regulating curcumin biosynthesis. The biosynthetic pathway for curcuminoid synthesis based on transcriptome, radiotracer and enzyme kinetic studies is represented in **Fig. 3**. Annadurai et al. (2013) constructed Illumina based rhizome specific transcriptomes of turmeric and the unigenes showing similarity to curcumin synthases were identified.

Sheeja et al. (2015) constructed rhizome specific transcriptomes from a high curcumin turmeric variety (Megha turmeric) and wild turmeric (*C. aromatica*) with negligible amount of curcumin using Illumina HiSeq 2000 platform. Differential gene expression analysis showed that all the ten putative curcumin biosynthetic pathway genes

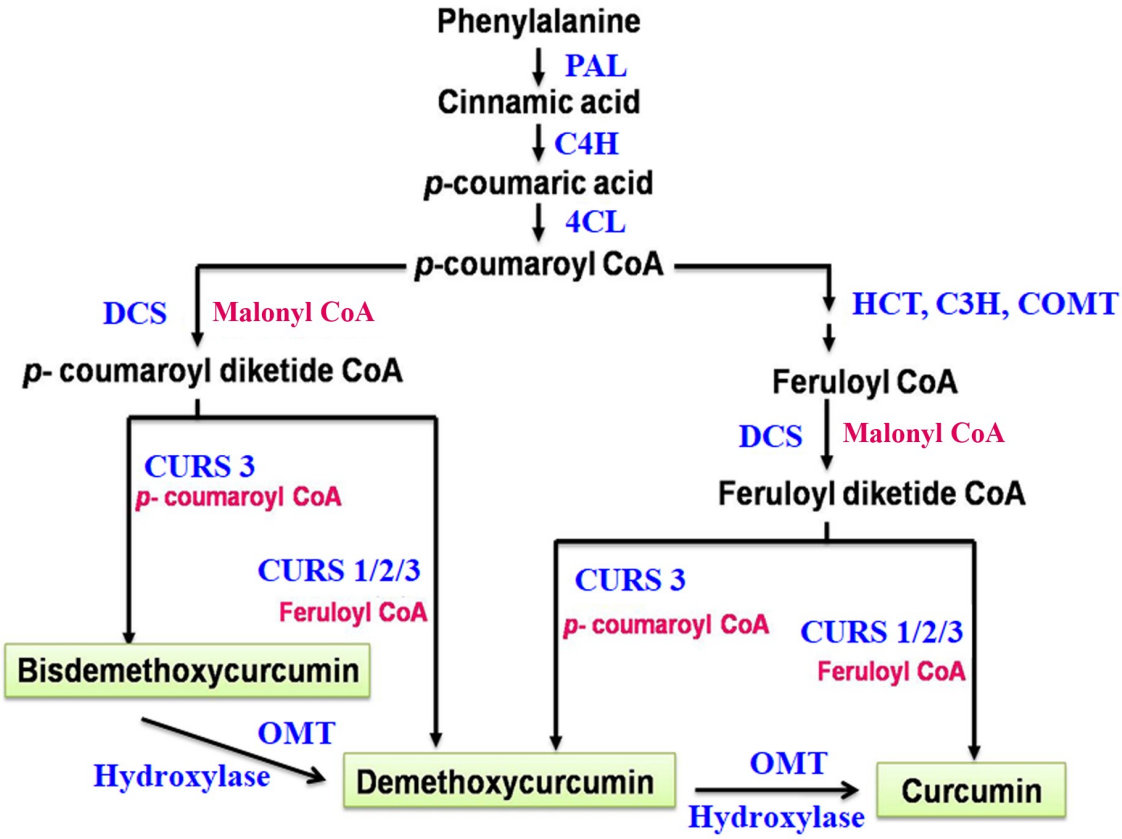


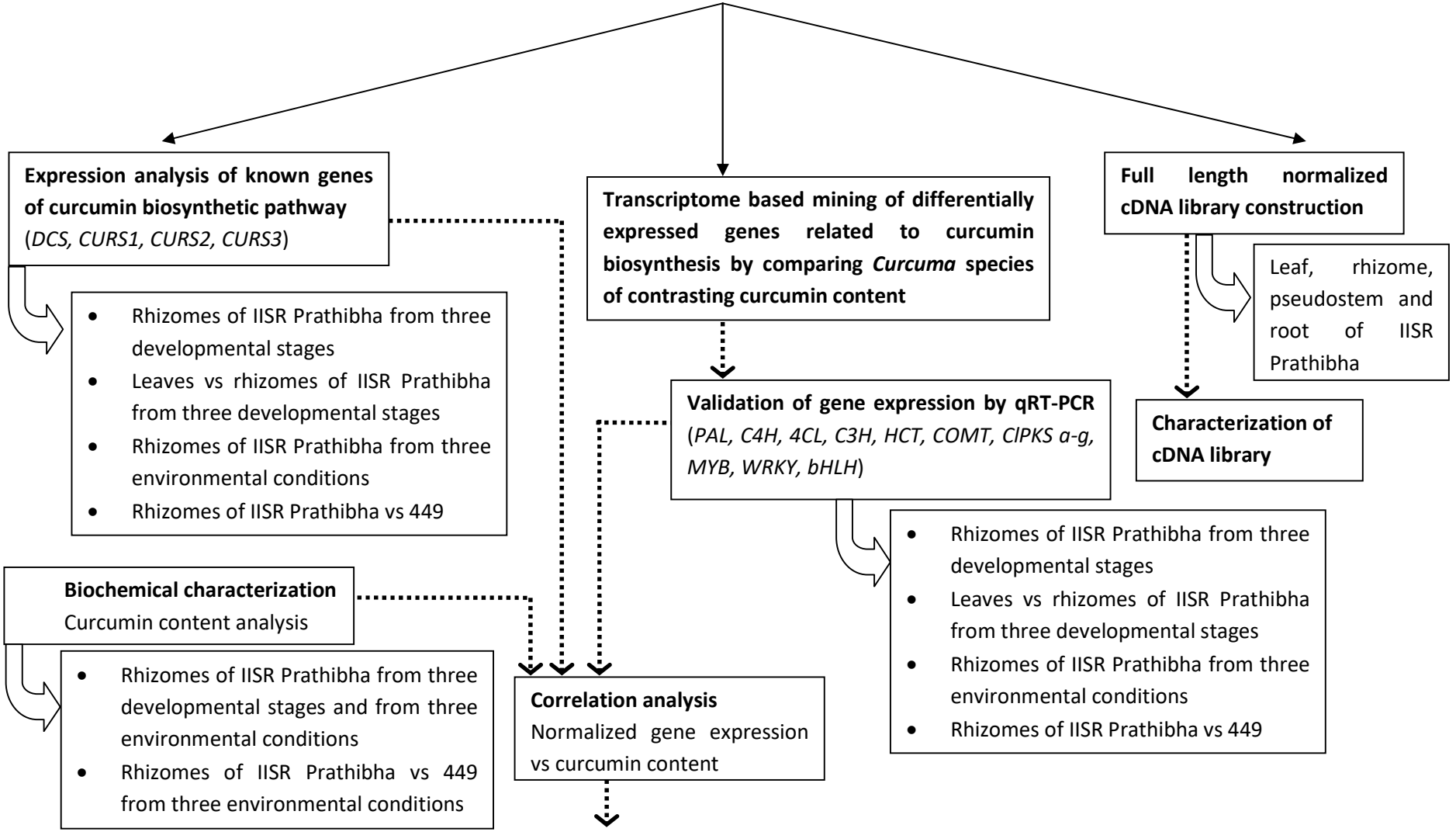
Figure 3. Schematic representation of curcuminoid biosynthesis in turmeric. PAL-Phenylalanine ammonia lyase, C4H-Cinnamate 4-hydroxylase, 4CL-4-Coumarate-CoA ligase, HCT-Hydroxy cinnamoyl transferase, C3H-Cinnamte 3-hydroxylase, COMT-Caffeoyl CoA O-methyltransferase, DCS-Diketide CoA synthase, CURS-Curcumin synthase, OMT-O-Methyltransferase

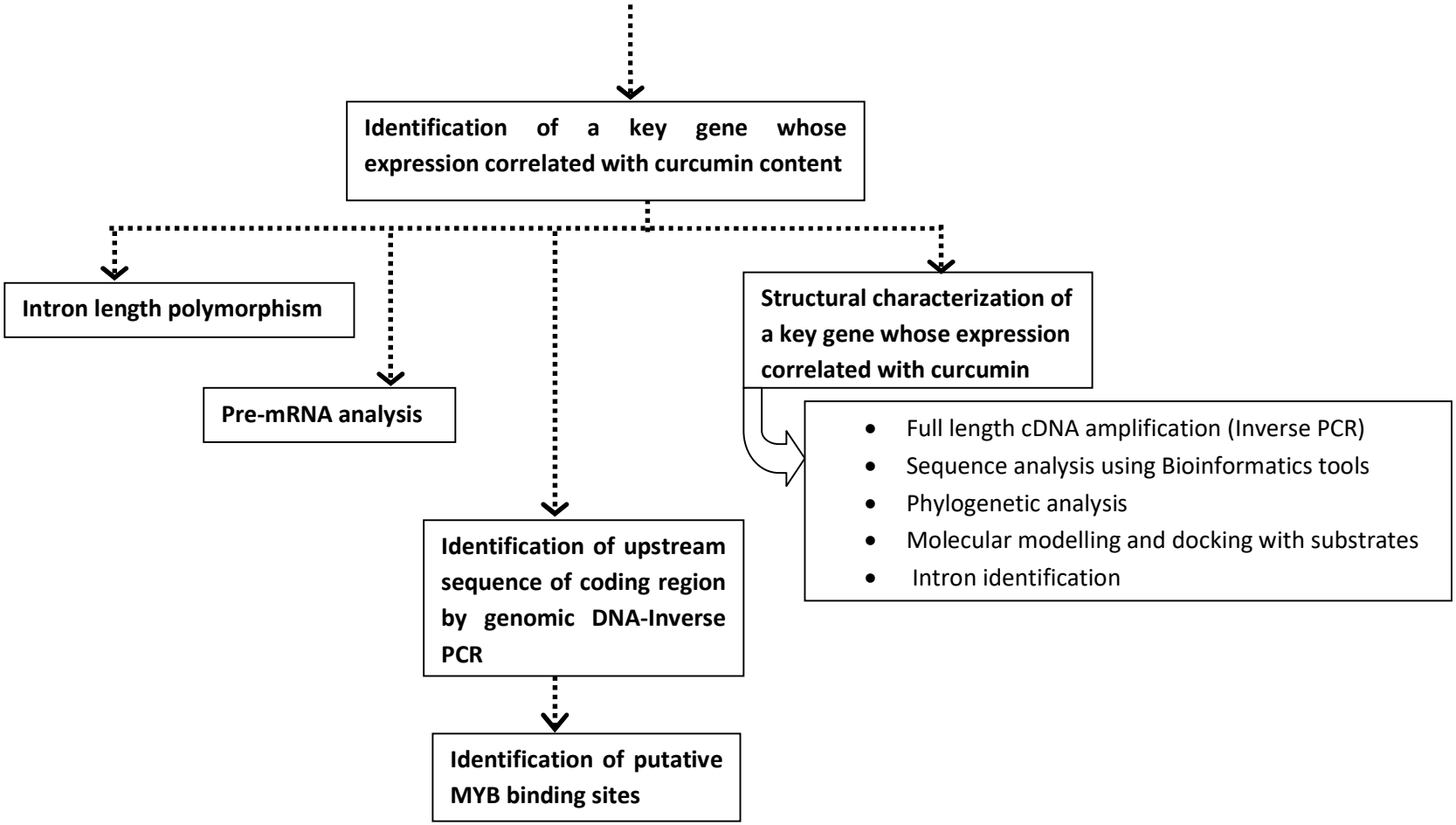
were present in both the transcriptomes with no significant variation in their expression level. Two novel polyketide synthases were identified with higher expression in Megha turmeric. The unigenes showing identity to 39 transcription factor families of AGRIS database were identified with AP2-EREBP and MYB as the abundant families. The validation of gene expression by real time PCR analysis of ten putative curcumin biosynthetic pathway genes, two novel polyketide synthases and two transcription factors (WRKY and MYB) was carried out in the study and was found to be on par with the results of transcriptome based differential gene expression analysis.

Wang (2016) constructed a complete biosynthetic pathway for synthesis of bisdemethoxycurcumin and curcumin in *E. coli* from an amino acid precursor, tyrosine. The co-expression of tyrosine ammonia lyase (TAL), 4-coumarate: CoA ligase (4CL) and curcuminoid synthase (CUS) from *Oryza sativa* yielded bisdemethoxycurcumin with a titre value of 28.9 ± 1.0 mg/l. Also the co-expression of 4-coumarate 3-hydroxylase (C3H) and catechol-O-methyltransferase (COMT) along with TAL, 4CL and CUS resulted in curcumin formation with a titre value of 0.6 ± 0.1 mg/l.

Methodology...

Transcriptomic approach to identify the genes involved in curcumin biosynthesis




Identification of a key gene whose expression correlated with curcumin content
Intron length polymorphism
Pre-mRNA analysis
Identification of upstream sequence of coding region by genomic DNA-Inverse PCR
Identification of putative MYB binding sites
Structural characterization of a key gene whose expression correlated with curcumin
Full length cDNA amplification (Inverse PCR)
Sequence analysis using Bioinformatics tools
Phylogenetic analysis
Molecular modelling and docking with substrates
Intron identification

Methodology

3.1 Plant material

IISR Prathibha, a turmeric variety released from Indian Council of Agricultural Research-Indian Institute of Spices Research (ICAR-IISR) with high curcumin content was used to optimize the RNA isolation protocol and to isolate genomic DNA. For normalized cDNA library construction, another high curcumin turmeric variety, Megha turmeric was utilized. These plants were maintained in greenhouse at ICAR-IISR, Kozhikode. The rhizomes of three genotypes of *C. longa* (Collection no. 19, RH17 and Megha turmeric), three different species of *Curcuma* (*C. xanthorrhiza*, *C. amada* and *C. caesia*) and ginger (*Zingiber officinale*) maintained at ICAR-IISR experimental farm, Peruvannamuzhi were used to confirm the utility of optimized RNA isolation protocol. To study gene expression analysis correlated with curcumin content and for proteomic analysis, IISR Prathibha and Collection no. 449 (low curcumin genotype) were planted under three different environments (**Fig. 4a-e**) namely Kozhikode-field, Kozhikode-green house and Coimbatore-field. To study the diversity in sequence of key curcuminoid biosynthetic pathway genes, three low curcumin genotypes (Collection no. 200, Collection no. 19 and Collection no. 449), three high curcumin genotypes (Megha turmeric, IISR Alleppey supreme and IISR Prathibha) and three zero curcumin *Curcuma* spp. (*C. aromatica, C. amada* and *C. caesia*) maintained at Kozhikode-field were used (**Fig. 4f**).

3.2 Details of molecular weight markers and cloning vector used

100 bp and 1 kb DNA ladders (**Fig. 5a, 5b**) purchased from Thermo Scientific and 200 bp DNA marker (**Fig. 5c**) from Takara were used to compare the size of the PCR products. Protein molecular weight marker (**Fig. 5d**) from Takara was used to compare the size of the peptides. For the construction of recombinants, TA cloning vector (pGEM-T Easy vector, Promega) was used (**Fig. 6**).

3.3 RNA isolation from different tissues of *Curcuma longa*

3.3.1 Precautions taken to avoid RNA degradation

a. All the glass wares and plastic wares were autoclaved before use

b. Water used for reagent preparation and rinsing the glass wares was treated with 0.1% diethylpyrocarbonate (DEPC)

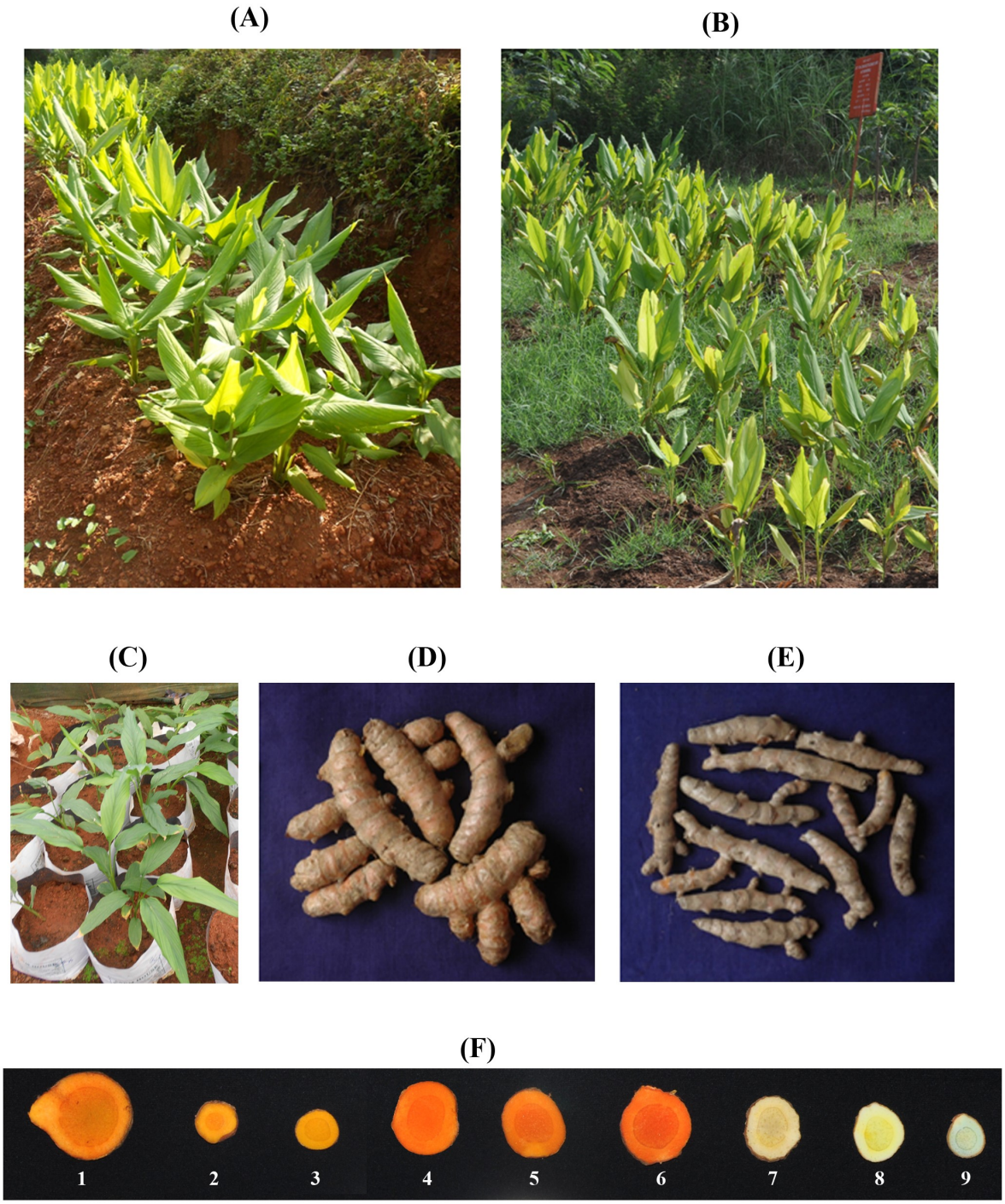


Figure 4. View of turmeric seedlings from (A) Kozhikode-field, (B) Coimbatore-field, (C) Kozhikode-green house, (D) Rhizomes of IISR Prathibha and (E) Collection no. 449; (F) Cross-section of primary rhizomes from three low curcumin turmeric collections [(1) Collection no. 200, (2) Collection no. 19 and (3) Collection no. 449], three high curcumin turmeric varieties[(4) Megha turmeric, (5) IISR Alleppey supreme and (6) IISR Prathibha] and three zero curcumin genotypes of *Curcuma* [(7) *C. aromatica,* (8) *C. amada* and (9) *C. caesia*]

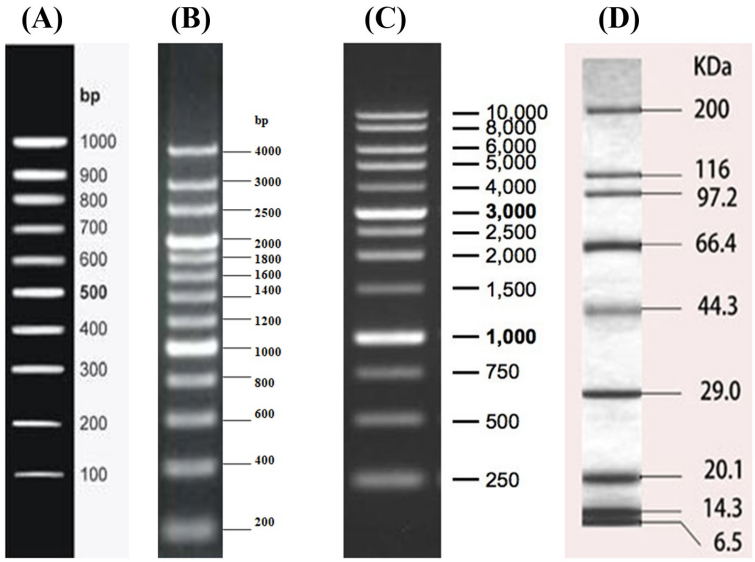


Figure 5. Map of different DNA rulers and protein molecular weight marker. 100 bp DNA ladder (A), 200 bp DNA ladder (B), 1 kb DNA ladder (C) and protein molecular weight marker (D)

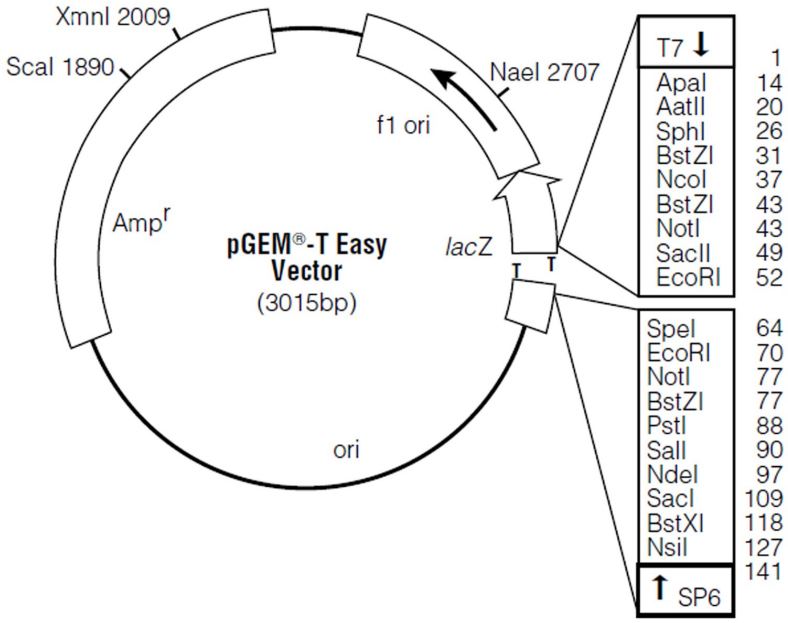


Figure 6. Map of pGEM-T Easy vector

c. The work bench was wiped with RNase Zap (Invitrogen) to remove the RNase contamination

d. Disposable gloves were used throughout the procedure to prevent the RNAse contamination from skin

e. The tissue was washed thoroughly and immediately processed either by freezing using liquid nitrogen or by storing in RNAlater (Invitrogen)

f. All the steps were carried out at 4ºC using refrigerated centrifuge and during incubation; the tubes were kept in ice unless otherwise mentioned

Five protocols namely *viz*., Trizol (Chomczynski and Sacchi, 1987) method, Plant RNeasy kit (Qiagen), CTAB method (Chang et al., 1993), SDS-Tris saturated phenol method (Ghawana et al., 2011) and SDS-acid phenol method (Hou et al., 2011) were carried out to isolate RNA from turmeric tissues. All the tissues were ground to a fine powder using liquid nitrogen.

3.3.2 Trizol method (Chomczynski and Sacchi, 1987)

One millilitre of Trizol (Invitrogen) reagent was added to powdered tissue (50 mg/100 mg) in a tube and vortexed thoroughly. The tubes were incubated for 5 minutes at room temperature (20-25ºC) for complete dissociation of nucleoprotein complex. The tubes were centrifuged at 12000 g for 10 minutes at 4ºC and the supernatant was transferred to a new tube. Subsequently, 0.2 ml of ice cold chloroform was added to the supernatant, mixed by inversion and incubated for 2-3 minutes at room temperature (20-25ºC). The samples were centrifuged at 12000 g for 15 minutes at 4ºC and the supernatant was transferred to a new tube. About 0.5 ml of ice cold isopropanol was added and incubated at room temperature (20-25ºC) for 10 minutes. The tubes were centrifuged at 12000 g for 10 minutes at 4ºC and the supernatant was discarded. One millilitre of 75% ethanol was added to the pellet, vortexed briefly and centrifuged at 7500 g for 5 minutes at 4ºC. The supernatant was discarded and the pellet was air dried for 5-10 minutes at room temperature (20-25ºC). The pellet was gently re-suspended in 25 µl RNase free water and incubated at 60ºC for 15 minutes.

3.3.3 Plant RNeasy kit (Qiagen) method

450 ml of buffer RLC and 4.5 µl of β-mercaptoethanol were added to 100 mg of powdered tissue taken in a tube and vortexed vigorously. The lysate was transferred to

QIA shredder spin column placed in 2 ml collection tube and centrifuged for 2 minutes at 12000 g at 4ºC. The supernatant was transferred to a new tube and 0.5 volume of 100% ethanol was added. The sample was transferred to RNeasy mini spin column placed in collection tube and centrifuged at 12000 g for 30 seconds at 4ºC. The flow through was discarded and 700 µl of Buffer RW1 was added to the RNeasy spin column and centrifuged at 12000 g for 30 seconds at 4ºC. The flow through was discarded and 500 µl of Buffer RPE was added to the RNeasy spin column and centrifuged at 12000 g for 30 seconds at 4ºC. The flow through was discarded and 500 µl of Buffer RPE was added to the RNeasy spin column and centrifuged at 12000 g for 2 minutes at 4ºC. The RNeasy spin column was placed in a new 2 ml collection tube and centrifuged at 12000 g for 1 minute at 4ºC to dry the membrane. The RNeasy spin column was transferred to a new 1.5 ml centrifuge tube, 30 µl of RNase free water was added to spin column membrane and centrifuged at 12000 g for 1 minute at 4ºC.

3.3.4 CTAB method (Chang et al., 1993)

One millilitre of extraction buffer (2% CTAB, 2% PVP, 0.1 M Tris HCl-pH8, 0.025 M EDTA-pH8, 2 M NaCl and 2% β-mercaptoethanol) pre-warmed at 65ºC was added to 100 mg of powdered tissue and vortexed thoroughly. Equal volume of chloroform: isoamyl alcohol (24:1) was added and centrifuged at 10000 g for 10 minutes at 4ºC. The supernatant was transferred to a new tube and an equal volume of chloroform: isoamyl alcohol (24:1) was added and centrifuged at 10000 g for 10 minutes at 4ºC. The supernatant was transferred to a new tube and 0.25 volumes of 10 M LiCl was added and incubated overnight at 4ºC. The tubes were then centrifuged at 10000 g for 20 minutes at 4ºC and the supernatant was discarded. The pellet was dissolved in 0.5 ml of SSTE (1 M NaCl, 0.5% SDS, 0.01 M Tris HCl-pH8, 0.001 M EDTA-pH8). 0.5 ml of chloroform: isoamyl alcohol (24:1) was added, mixed well and centrifuged at 10000 g for 10 minutes at 4ºC. The supernatant was transferred to a new tube and 2 volumes of 100% ethanol were added. The tubes were incubated at -70ºC for 30 minutes and centrifuged at 10000 g for 20 minutes at 4ºC. The supernatant was discarded and the pellet was air dried for 5-10 minutes at room temperature (20-25ºC). The pellet was gently re-suspended in 25µl RNase-free water.

3.3.5 SDS-Tris saturated phenol method (Ghawana et al., 2011)

Two millilitre of extraction buffer (0.1% SDS, 0.32 M sodium acetate, 0.01 M EDTA-pH 8 in phenol saturated with 0.01 M Tris HCl-pH 8 and 0.001 M EDTA) was taken in a pre-chilled mortar and 100 mg of powdered tissue was added and ground thoroughly using a pestle till it thawed completely. DEPC-treated water (0.8 ml) was added and mixed by grinding. The lysate was transferred to two, 2 ml centrifuge tubes and incubated at room temperature (20-25ºC) for 5 minutes. Chloroform (0.2 ml) was added to each tube, vortexed and incubated at room temperature (20-25ºC) for 10 minutes. The tubes were centrifuged at 13000 g for 10 minutes at 4ºC and the aqueous phase was transferred to a fresh tube. About 0.6 volumes of ice cold isopropanol was added to the aqueous phase, vortexed, incubated at room temperature (20-25ºC) for 10 minutes and centrifuged at 13000 g for 10 minutes at 4ºC. One millilitre of 75% ethanol was added to the pellet, vortexed briefly and centrifuged at 7500 g for 5 minutes at 4ºC. The supernatant was discarded and the pellet was air dried for 5-10 minutes at room temperature (20-25ºC). The pellet was gently re-suspended in 25 µl RNase free water.

3.3.6 SDS-acid phenol method (Hou et al., 2011)

Extraction buffer (0.1 M Tris HCl, 0.4 M NaCl, 1% SDS, 2% β- mercaptoethanol) of 0.6 ml was taken in a 1.5 ml centrifuge tube and 500 mg of powdered tissue was added. The mixture was vortexed thoroughly and incubated for 3 minutes. A volume of 0.4 ml of acid phenol: chloroform-pH 4.5 (Invitrogen) was added, centrifuged at 10000 g for 10 minutes at 4ºC and the supernatant was transferred to a new tube. Five molar NaCl (0.2 ml) and 0.4 ml of acid phenol: chloroform-pH 4.5 (Invitrogen) was added to the supernatant, centrifuged at 10000 g for 10 minutes at 4ºC and the supernatant was transferred to a new tube. Equal volume of isopropanol was added to the supernatant, incubated at room temperature (20-25ºC) for 10 minutes and centrifuged at 10000 g for 10 minutes at 4ºC. One millilitre of 70% ethanol was used to wash the pellet and centrifuged at 10000 g for 10 minutes at 4ºC. The supernatant was discarded and the pellet was air dried for 5-10 minutes at room temperature (20-25ºC). The pellet was gently re-suspended in 25µl RNase-free water

3.3.7 Modified protocol for total RNA extraction

Hundred milligram of powdered tissue was added to 2 ml extraction buffer [SDS (0-3%), 0.1 M Tris HCl-pH 8 and 0.025 M EDTA-pH 8, PVP (0-3%) and β mercaptoethanol (0-3%) pre-warmed at 65˚C taken in a mortar and ground thoroughly. The lysate was transferred to two, 2 ml centrifuge tubes and equal volume of acid phenol: chloroform-pH 4.5 (Ambion) was added. The mixture was vortexed for 10 seconds, incubated at room temperature (20-25˚C) and centrifuged at 15000 g for 10 minutes at 4˚C. The supernatant was transferred to new tubes and 0.3 volume of (1-5 M) sodium acetate and 0.7 volume of acid phenol: chloroform-pH 4.5 was added, mixed by inversion, incubated on ice for 10 minutes and centrifuged at 15000 g for 10 minutes at 4ºC. The supernatant was transferred to new tubes. To the supernatant 0.1 volume of (1-5 M) sodium acetate and equal volume of isopropanol was added, incubated at -20˚C for 1h and centrifuged at 15000 g for 10 minutes at 4˚C. The supernatant was discarded; 1 ml of 70% ethanol was added to the pellet and centrifuged at 7500 g for 10 minutes at 4ºC. The supernatant was discarded; pellet was air dried for 5-10 minutes at room temperature (20-25ºC) and re-suspended in 25 µl RNase-free water.

3.3.8 Determination of concentration and purity of RNA

The RNA was quantified using Biophotometer plus (Eppendorf) with 0.2 mm lid using 2 µl of total RNA. The concentration (ng/µl), A_{260}/A_{280} ratio and A_{260}/A_{230} ratio of the sample were recorded. The integrity of RNA was checked by denaturing agarose gel electrophoresis (Masek et al., 2005). The precautions described in section 3.3.1 were followed to prevent RNA degradation. Agarose (1.2%) was prepared by adding 1.2 g of agarose in 100 ml of 1X TAE buffer prepared from 50 X TAE (2 M Tris-pH 8, 1 M glacial acetic acid and 50 mM EDTA-pH 8). The solution was boiled for 3 minutes and cooled to room temperature and 2 µl of ethidium bromide (10 mg/ml) was added, swirled gently and poured to the gel casting tray fitted in gel casting unit. The comb was placed and the gel was allowed to solidify for 20 minutes. The solidified gel was placed in the electrophoresis unit containing 1X TAE buffer and the comb was removed. Five microlitre of total RNA was mixed with 18 µl of formamide and 2.5 µl of 10X loading dye (0.05 M Tris-pH 7.6, 0.2% bromophenol blue and 60% glycerol) and denatured by incubating at 65ºC for 5 minutes. The tubes were snap cooled on ice for 5 minutes and resolved in 1.2% TAE agarose gel under constant current of 90 V till the bromophenol blue migrated to the

bottom of the gel. The gel was visualized under UV using gel documentation system (Syngene).

The RIN value was analyzed using Agilent RNA 6000 Nano kit on Agilent 2100 Bioanalyzer. The reagents were equilibrated to room temperature for 30 minutes prior to use. The chip priming station was arranged by placing new syringe into the hole of the luer lock adapter and screwed it tightly to the chip priming station. The base plate was adjusted to position C and the lever of the syringe clip was slided to the top position. The electrodes were decontaminated with RNAse Zap (Invitrogen) for 1 minute and with RNAse free water for 10 seconds. RNA ladder was denatured by heating it at 70ºC for 2 minutes followed by snap cooling in ice and aliquots were stored in -70ºC. The gel was filtered by passing 550 µl of Agilent RNA 6000 Nano gel matrix through a spin filter, centrifuged at 4000 rpm for 10 minutes and aliquots of 65 µl each was transferred to 0.5 ml tubes. To an aliquot of 65 µl filtered gel, 1 µl of RNA 6000 Nano dye concentrate was added, vortexed and centrifuged at 14000 rpm for 1 minute. Nine microlitre of the gel dye mix was loaded carefully into the bottom of the well marked 'G' on the Agilent Nano chip placed on the chip priming station. The chip priming station was closed making sure that the plunger was positioned at 1 ml. The plunger of the syringe was pressed down and after 30 seconds, the plunger was released. The chip priming station was opened and 9 µl of the gel-dye mix was transferred to the wells marked G. Five microlitre of the RNA 6000 Nano marker was loaded in all the 12 sample wells and in the ladder well. One microlitre each of the RNA sample was loaded into the sample wells and 1 µl of denatured RNA ladder into the ladder well. The chip was vortexed for 1 minute at 2400 rpm, inserted in Agilent 2100 Bioanalyzer and read.

3.4 Isolation and integrity analysis of low molecular weight (LMW) RNA

Low molecular weight RNA was isolated following Lu et al. (2007). One tenth volume of 50% PEG and 1/10th volume of 5 M NaCl was added to total RNA having concentration of 200-400 µg dissolved in RNase free water. The tubes were mixed by inversion, incubated on ice for 30 minutes and centrifuged at 12000 g for 10 minutes at 4˚C. The supernatant was transferred to a new tube and 2.5 volumes of 100% ethanol was added. The tubes were mixed by inversion, incubated at -20˚C for 2 hours and centrifuged at 12000 g for 30 minutes at 4˚C. The supernatant was removed by pipetting, 0.5 ml of 80% ethanol was added to the pellet and centrifuged at 7500 g for 10 minutes at 4ºC. The

supernatant was discarded by pipetting, pellet was air dried for 5-10 minutes at room temperature (20-25ºC) and re-suspended in 25 µl RNase-free water.

The integrity of small RNA was analysed by 15% denaturing PAGE as described by Lu et al. (2007). All the steps were carried out in RNAse free environment following the precautions as described in section 3.3.1. Polyacrylamide/urea gel (15%) was prepared by mixing 9.6 g urea, 7.5 ml of 40% acrylamide and 2 ml of 10X TBE in DEPC treated water to a final volume of 20 ml and the solution was warmed to dissolve the urea. After cooling to room temperature, 120 µl of freshly prepared 10% ammonium persulfate and 9.2 µl of TEMED was added to the acrylamide solution and mixed by swirling. The solution was poured to the assembled plates (10x8 cm) with spacers (1.5 mm). The comb was inserted immediately after pouring the solution. After 30 minutes, the comb was removed carefully and the wells were rinsed with 1X TBE. The gel was pre-run for 20 minutes at 200 V. Ten µl of low molecular weight RNA was mixed with equal volume of 2X loading dye (0.01% SDS, 0.01% bromophenol blue, 0.005% xylene cyanol, 0.05 M EDTA and 47.5% formamide). In another tube, 3 µl of 10 bp ladder was mixed with equal volume of 2X loading dye. Both the sample-dye mix and ladder-dye mix were loaded to the respective wells after denaturing at 65ºC for 5 minutes followed by snap cooling on ice for 2 minutes. The samples were resolved using PAGE with 1X TBE at 200 V for 1 h (till the bromophenol blue dye front migrated to the bottom of the gel). The gel was stained by soaking it in 100 ml of 0.5X TBE containing 10 µl SYBR gold for 5 minutes and incubated in dark for 10 minutes and washed with 0.5X TBE. The gel was visualized under UV using gel documentation system (Syngene).

3.5 DNase treatment and cDNA synthesis

The presence of DNA in the isolated RNA was removed using DNase I (Thermo Scientific). To an RNase free tube, 1 µg of RNA, 1 µl of reaction buffer (10X) with $MgCl_2$ and 1 µl of DNase (1 U/µl) was added and made up to 10 µl with RNase free water. The tube was incubated at 37ºC for 30 min. After the incubation, DNase was inactivated by adding 1 µL of 50 mM EDTA and incubated at 65°C for 10 min and then snap cooled on ice. DNase free RNA was reverse transcribed to cDNA using superscript III reverse transcriptase (Invitrogen). To a sterile tube, 250 ng of DNase treated RNA, 1 µl of 100 µM oligo$(dT)_{18}$ and 1 µl of 10 mM each dNTP mix was added and made up to 13 µl using RNase free water. The tube was incubated at 65ºC for 5 minutes and then chilled on ice. To

this RNA- oligo(dT)$_{18}$ –dNTP mix, 4 µl of 5X reaction buffer, 1 µl of 0.1 M DTT, 1 µl of Ribolock RNase inhibitor (40 U/µl) and 1 µl of 200 U superscript III reverse transcriptase was added. The reagents were mixed gently and incubated at 50ºC for 60 minutes. After incubation, the reaction was terminated by heating at 70ºC for 15 minutes and then cooled to 4ºC.

3.6 Amplification of gene of interest by polymerase chain reaction

The amplification of gene of interest was carried out by Polymerase Chain Reaction (PCR) with gene specific primers. The gene specific primers were designed using Primer3plus, an online primer designing software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) using default parameters and synthesized from Sigma. The primers in lyophilized form were suspended in sterile water to a stock concentration of 100 µM and diluted 10 times to the working concentration of 10 µM. Two microlitre of template DNA (25 ng/µl), 2.5 µl of 10X PCR buffer, 1 µl of dNTP mix (2.5 mM each), 1 µl each of forward and reverse primer (10 µM) and 0.1 µl of 5 U/µl Taq DNA polymerase (Takara) was added to a sterile 0.2 ml tube. The contents were made upto 25 µl with sterile water, mixed gently and placed in thermal cycler (Eppendorf). The PCR cycle comprised an initial denaturation for 2 minutes at 95ºC, 30 cycles of denaturation at 94ºC for 45 seconds, annealing temperature depending on the primers for 45 seconds, extension at 72ºC for 1 minute/kb and final extension at 72ºC for 10 minutes. The products were stored at 4ºC. The amplified PCR product was mixed with 5 µl of 6X gel loading dye (0.25% bromophenol blue, 0.25% xylene cyanol and 30% glycerol in sterile water) and resolved in 1% agarose gel under constant current of 90V till the bromophenol blue dye front migrated to the bottom of the gel. The gel was visualized under UV using gel documentation system (Syngene).

3.7 Quantitative real time PCR

To check, the quality of RNA, stability of reference genes and relative gene expression analysis; quantitative real time PCR (qRT PCR) was carried out. The reaction was done in Rotor Gene Q (Qiagen) with 72 well rotor using QuantiFast SYBR Green PCR kit (Qiagen). The pooled RNA from three biological replicates, treated with DNase and reverse transcribed was used as the template. All the samples to be analyzed for a particular gene was processed and run together for accurate relative gene expression

analysis. The gene specific primers for qRT PCR were designed using Primerquest, an online primer designing software (https://eu.idtdna.com/Primerquest). The parameters used for primer designing were 100-150 bp amplicon length, 60-62^{o}C melting temperature and 40-60% GC content. The primers were synthesized from Sigma. The primers in lyophilized form were suspended in sterile water to a stock concentration of 100 µM and diluted 10 times to the working concentration of 10 µM.

The reaction mixture comprised of 10 µl of 2X SYBR green, 1 µl each of 10 µM gene specific forward and reverse primer, 0.4 µl of cDNA and made upto a final volume of 20 µl with sterile water. Three technical replicates of each reaction were performed. A reverse transcription negative control (without reverse transcriptase) and a non-template negative control were included to confirm the absence of genomic DNA and to check non-specific amplification respectively. PCR amplification was carried out with initial denaturation at 95ºC for 5 minutes followed by 35 cycles of 95ºC for 10 seconds and 60ºC for 30 seconds. A melt curve program was done by increasing the temperature stepwise by 1^{o}C in the range of 65-99^{o}C to check the specificity of PCR products. Mean Cq values of three technical replicates was exported after setting the threshold to 0.1 and correcting the slope and baseline. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) and transformed to $\log_2$ scale. A $\log_2$-ratio of zero indicates the transcript has equal expression values, whereas the transcripts are up-regulated if the ratio is above zero and down-regulated if the ratio is below zero (Khraiwesh et al., 2015). A $\log_2$ ratio of 1 represents a 2-fold change and the transcripts with fold change ≥ 1 / ≤ -1 were defined as significantly up-/down-regulated (Luo et al., 2005). The PCR products were run on 1.2% agarose gel and sequenced to confirm its specificity.

3.8 Construction of normalized cDNA library

3.8.1 First strand cDNA synthesis

The cDNA was synthesized using SMARTer PCR cDNA synthesis kit (Clontech). The mechanism involved is schematically represented in **Fig. 7**. To a sterile tube, 1 µg of pooled RNA comprising 250 ng each from leaves, pseudostem, rhizomes and root tissues was added and denatured at 65ºC for 2 minutes. The tube was spun gently. To the RNA, 1 µl of SMARTer IIA oligonucleotide and 1 µl of 3' SMART CDS primer IIA was added and made up to 5 µl with sterile water. The reaction mixture was gently mixed by pipetting

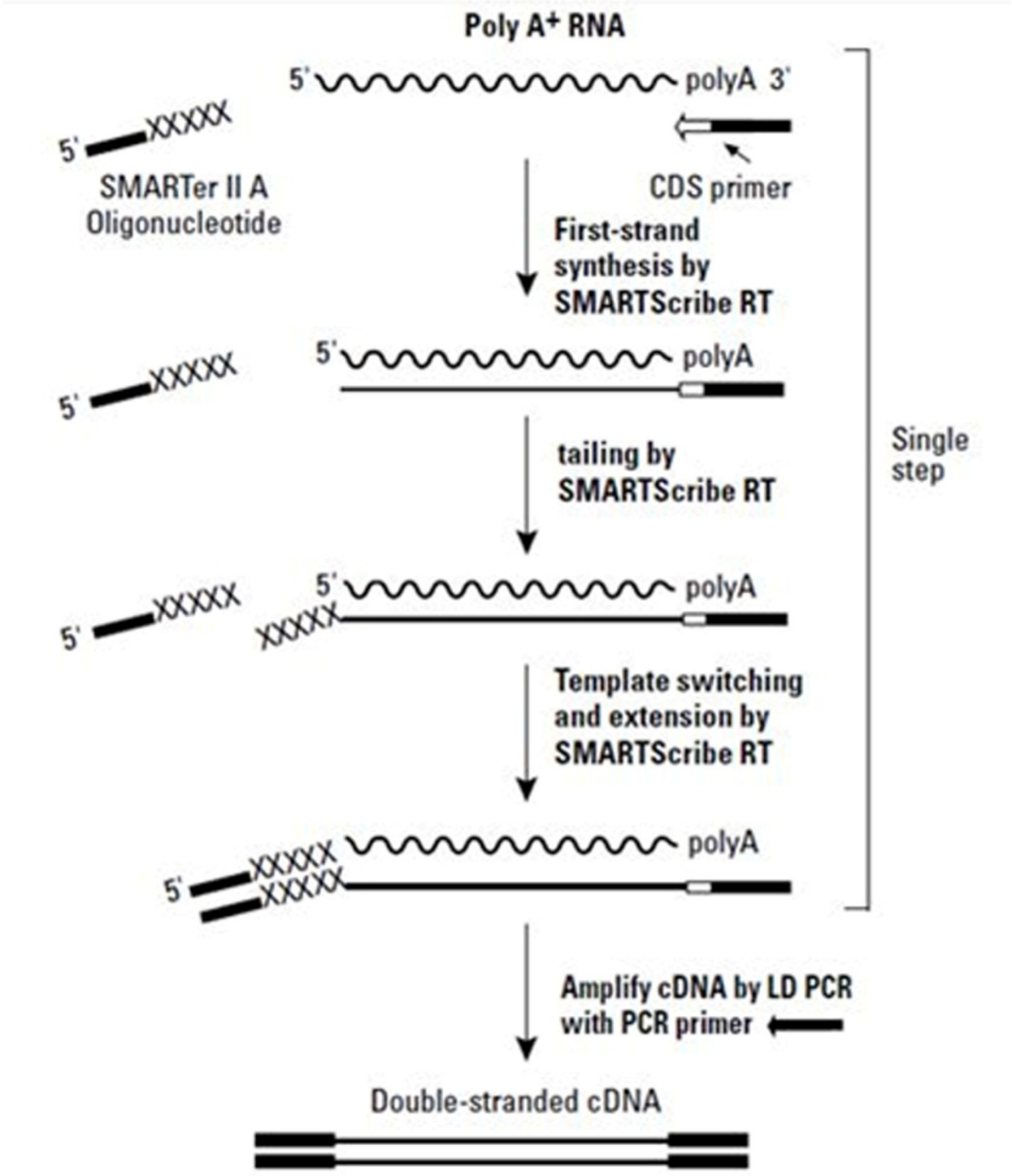


Figure 7. Schematic diagram of ds cDNA synthesis (SMARTer PCR cDNA synthesis kit user manual, Clontech)

and incubated at 70ºC for 2 minutes followed by 42ºC for 2 minutes. After incubation, 2 µl of 5X first strand buffer, 0.2 µl of 0.1 M DTT, 1 µl of 50X dNTP mix, 0.25 µl of 40 U/µl RNase inhibitor and 1 µl of 100 U/µl SMARTScribe MMLV reverse transcriptase were added, mixed by pipetting and incubated at 42ºC for 90 minutes and snap cooled on ice to terminate the first strand cDNA synthesis. A positive control reaction was carried out with 1 µg of mouse liver total RNA provided in the kit.

3.8.2 Double stranded cDNA synthesis

To an aliquot of 2 µl first strand cDNA, 10 µl of 10X advantage 2 PCR buffer, 2 µl of 10 mM each of dNTP mix, 4 µl of 10 µM PCR primer M1 and 2 µl of 50X advantage 2 polymerase mix and 80 µl sterile water was added, mixed by pipetting and subjected to PCR cycling using the following program. An initial denaturation of 95ºC for 1 minute followed by 15-18 cycles of 95ºC for 15 seconds, 66ºC for 20 seconds and 72ºC for 3 minutes was carried out. The PCR products were resolved in 1.5% agarose gel and documented.

3.8.3 PCR product purification

Double stranded cDNA was purified using QIA quick PCR purification kit (Qiagen). Five volumes of buffer PB was added to the PCR product and mixed by inversion. The contents were transferred to QIA quick spin column placed on 2 ml collection tube and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded, 750 µl of buffer PE was added to the spin column and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and a dry spin was done by centrifuging the spin column at 13000 rpm for 1 minute. The spin column was placed in a sterile 1.5 ml centrifuge tube, 30 µl of sterile water was added to the centre of the column and centrifuged at 13000 rpm for 1 minute.

3.8.4 Normalization of ds cDNA

The full length transcripts were normalized using Duplex Specific Nuclease (DSN) and the major steps involved is represented in **Fig. 8**. One µg of purified ds cDNA and 4 µl of 4X hybridization buffer were made upto 16 µl using sterile water and an aliquot of 4 µl was transferred to four tubes. The samples were incubated at 98ºC for 2 minutes followed by 68ºC for 5 minutes. Five µl of 2 X DSN master buffers pre-heated at 65ºC for 5 minutes was added to the ds cDNA-hybridization buffer mix and incubated at 68ºC for 10 minutes. After incubation 1 U, 0.5 U and 0.25 U of DSN was added to three individual tubes and 1 µl of DSN storage buffer was added to another tube as negative control and incubated at 68ºC for 25 minutes. After the incubation, 5 µl of DSN stop solution was added and incubated at 68ºC for 5 minutes and snap cooled on ice followed by the addition of 25 µl of sterile water to each tube.

3.8.5 Amplification of normalized ds cDNA

One microlitre of cDNA treated with different concentrations of DSN (1U, 0.5U and 0.25U) and negative control (ds cDNA without DSN) was aliquoted to separate sterile tubes. Five microlitre of 10X advantage 2 PCR buffer, 10 mM each of 50X dNTP mix, 1.5 µl of 10 µM PCR primer M1 and 1 µl of 50X advantage 2 polymerase mix was added to the sample, made upto 50 µl with sterile water and subjected to PCR cycling. An initial denaturation at 95ºC for 1 minute followed by 7-13 cycles of denaturation at 95ºC for 15 seconds, primer annealing at 66ºC for 20 seconds and extension at 72ºC for 3 minutes was carried out. The optimization of number of PCR cycles was carried out with control tube by analyzing aliquots of 5 µl after 7^{th}, 9^{th}, 11^{th} and 13^{th} PCR cycles in 1.5% agarose gel. After the optimization, the sample tubes with different concentration of DSN was subjected to additional number of PCR cycles to reach the optimal number as determined using the control tube. The PCR products were resolved in 1.5% agarose gel and documented.

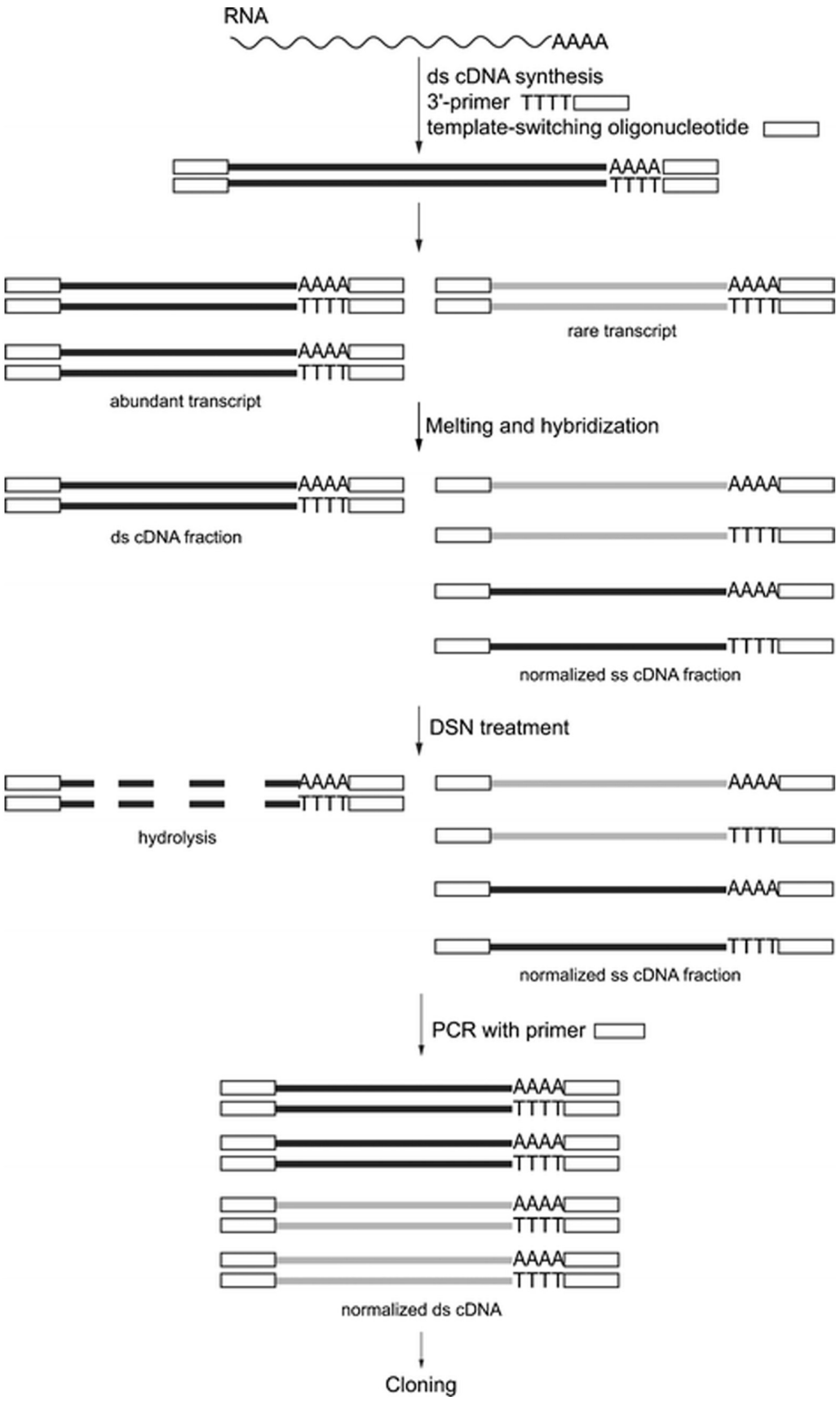


Figure 8. Outline of Duplex Specific Nuclease normalization (Bogdanova et al., 2008)

3.8.6 Non-directional cloning of ds cDNA

The normalized full length ds cDNA was purified using QIA quick PCR purification kit as mentioned in section 3.8.3. For ligation, 3 µl aliquot of the purified normalized full length ds cDNA, 5 µl of 2X rapid ligation buffer, 1 µl of 50 ng/µl pGEM-T Easy vector (Promega) and 1 µl of T4 DNA ligase was added. The sample was mixed by pipetting and incubated at 22ºC for 1 hour followed by overnight incubation at 4ºC. A control reaction was carried out with 8 ng of 0.5 kb PCR fragment provided in the kit. Meanwhile, a single isolated colony of *E. coli* JM109 was streaked on LB agar using quadrant streak method and incubated overnight at 37ºC for competent cell preparation. After the incubation, 1-2 colonies of 2x2 mm size was inoculated to 10 ml LB broth aseptically and incubated in a rotary shaker at 150 rpm at 37ºC for 3-4 h till the OD_{600} reached the range of 0.40-0.55. The culture was centrifuged at 6000 rpm for 3 minutes at 4ºC. The supernatant was discarded; the pellet was suspended in 5 ml ice cold 0.1 M $CaCl_2$ solution and incubated on ice for 45 minutes followed by centrifugation at 6000 rpm for 3 minutes at 4ºC. The supernatant was discarded and the pellet was suspended in 0.5 ml of ice cold 0.1 M $CaCl_2$ solution and 100 µl each was aliquoted into 0.5 ml tubes. To an aliquot of *E. coli* competent cells, 2 µl of ligated mixture was added and incubated on ice for 30 minutes followed by incubation at 42ºC for 90 seconds and then snap cooled on ice. About 350 µl of LB broth was then added and incubated at 37ºC for 40 minutes. After the incubation, the tubes were centrifuged at 8000 rpm for 1 minute. The supernatant was discarded and the pellet was resuspended in 50 µl of supernatant solution and plated on to the LB-ampicillin XGal/IPTG plates and incubated overnight at 37°C. pUC18 (100 ng) was used as the control to check the efficiency of competent cells. To prepare LB-ampicillin XGal/IPTG plates, 100 ml of LB agar was melted and cooled to 55ºC and 100 µl of 50 mg/ml ampicillin, 200 µl of 20 mg/ml X-Gal and 200 µl of 100 mM IPTG was added, swirled gently and allowed to solidify in sterile petriplates.

3.8.7 Confirmation of recombinants

The recombinants were confirmed using colony PCR and plasmid DNA sequencing. The white recombinant clones were back-streaked to fresh LB-ampicillin XGal/IPTG plates and incubated overnight at 37°C. To a sterile tube, 10 µl of sterile water was added and a small portion of the colony was transferred using sterile tooth pick, mixed thoroughly and briefly centrifuged. One microlitre of the supernatant was taken as the

template and PCR amplified as described in section 3.6 with M13 primers. The amplified PCR product was mixed with 5 µl of 6X gel loading dye and resolved in 1% agarose gel and documented.

The plasmid DNA was isolated following alkaline lysis method (Bimboim and Doly, 1979). The selected positive colonies were inoculated into 3 ml of LB broth with ampicillin (50 µg/ml) and incubated overnight in a rotary shaker at 150 rpm at 37ºC. The culture was centrifuged at 13000 rpm for 1 minute at 4ºC and the pellet was resuspended in 100 µl of ice-cold solution I (50 mM glucose, 25 mM Tris HCl-pH 8, 10 mM EDTA-pH 8 and 10 µg/ml RNase) by pipetting. To the lysed solution, 200 µl of freshly prepared solution II (0.2 N NaOH and 1% SDS) was added and mixed by inverting the tube 4-6 times and incubated on ice for 5 minutes. One hundred and fifty mirolitre of ice-cold solution III (3 M potassium acetate-pH 6, prepared by adding 60 ml of 5 M potassium acetate, 11.5 ml of glacial acetic acid and 28.5 ml sterile water) was added to the lysate mix, mixed thoroughly 4-6 times by inversion and centrifuged at 12000 rpm for 10 minutes at 4ºC. The clear lystate was transferred to a spin column, centrifuged at 10000 rpm for 1 minute at 4ºC and the flow-through was discarded. The spin column was then washed twice with wash solution (1:5 diluted solution of 50 mM Tris HCl-pH 8 with 100% ethanol) by centrifuging at 10000 rpm for 1 minute at 4ºC. The flow-through was discarded and a dry spin was done by centrifuging at 12000 rpm for 2 minutes at 4ºC. The spin column was transferred to a new 1.5 ml centrifuge tube. Nuclease free water (50 µl) was added to the centre of the column and centrifuged at 12000 rpm for 1 minute at 4ºC. The presence of plasmid DNA was checked by resolving 5 µl of plasmid DNA in 0.8% agarose gel and the gel was documented. The sequencing of recombinants was performed at Xcelris Labs Limited, Ahmedabad using ABI 3730xl 96 capillary system with Big Dye Terminator v3.1 kit. The vector sequences were removed using NCBI tool Vecscreen (http://www.ncbi.nlm. nih.gov/tools/vecscreen/) and their identity was compared in NCBI database using BLASTX (https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastx&PAGETYPE=BlastSearch&LINK_LOC=blasthome).

3.9 Transcriptome analysis

The annotated transcriptomes of *C. longa* and *C. aromatica* (Sheeja et al., 2015) were used to mine the genes related to curcumin biosynthesis. The differential gene expression analysis was carried out using DESeq program. Keyword search analysis in the

annotated files of *C. longa* transcriptome and local blast analysis in the Bioedit software (Hall, 1999) were carried out to identify the genes related to curcumin biosynthesis. The transcripts encoding transcription factors were identified by aligning against Arabidopsis Gene Regulatory Information Server (AGRIS) database and only the transcripts with >80% identity was reported.

3.10 Stability analysis of reference genes

The standard curves were generated for each reference gene with serial dilutions of pooled cDNAs (10^{-1} to 10^{-6}) as template by qRT-PCR and the reaction was carried out as mentioned in section 3.7. The PCR efficiency and regression coefficient value of each primer pair was calculated using the Rotor Gene Q software (Qiagen). The gene expression values of six reference genes namely; *EF1α, UBIQUITIN, 18S rRNA, GAPDH, ACTIN* and *TUBULIN* was analyzed in leaves, rhizomes, pseudostem and roots of 60, 120 and 180 DAP (Days After Planting) and exported to Reffinder (Xie et al., 2012) to evaluate the stability of reference genes.

3.11 Quantification of curcumin

The curcumin content was analyzed as per American Spice Trade Association (ASTA) procedure, 1968. One hundred milligram of powdered and sieved dried primary rhizomes was taken in a round bottom flask; 30 ml of ethanol was added and refluxed for 2.5 hrs. After the solution was cooled, it was filtered and made up to 100 ml with ethanol in volumetric flask. Two millilitre aliquot of the extract was made upto 25 ml in volumetric flask and the absorbance was measured at 425 nm against the ethanol as blank. A standard curcumin solution was prepared using curcumin (Sigma) with a final concentration of 0.0025 g/l. The percentage of curcumin in the sample was calculated as follows:

% of curcumin in the sample = (Absorbance of extract at 425 nm x 125) / (cell length in cm x absorptivity of curcumin x sample weight in grams)

Absorptivity of curcumin (α) = Absorbance of standard solution at 425 nm / (cell length in cm x concentration in g/l)

3.12 Correlation analysis

The relationship between normalized expression levels of the genes namely, ten reported candidate genes of curcumin biosynthesis (*PAL, C4H, 4CL, C3H, HCT, COMT, DCS, CURS1, CURS2* and *CURS3*) and eleven transcripts selected based on differential gene expression analysis, which includes seven transcripts orthologous to polyketide synthases, two MYBs, WRKY and bHLH vs. curcumin (%) analyzed from turmeric genotypes with contrasting curcumin content at different developmental stages and environmental conditions was performed using Pearson's correlation coefficient in MS-Excel.

3.13 Amplification of unknown sequence of target gene

3.13.1 Inverse PCR

Inverse PCR combined with SMARTer PCR cDNA synthesis as described by Kuniyoshi et al. (2006) was applied to amplify full length cDNA of target gene (**Fig. 9**). The first strand cDNA was synthesized from 1 µg of pooled RNA, comprising 250 ng each from leaves, rhizomes, pseudostem and roots of Megha turmeric harvested at 120 DAP. To the pooled RNA, 1 µl of 12 µM 3' SMART CDS primer IIA was added, made up to 4.5 µl with sterile water and incubated at 72°C for 3 minutes and then at 42°C for 2 minutes. After the incubation, 2 µl of 5X first strand buffer, 0.25 µl of 100 mM DTT, 1 µl of 10 mM dNTP mix, 1 µl of 12 µM SMARTer oligonucleotide, 0.25 µl of 40 U/µl Ribolock and 1 µl of 100 U SMARTScribe reverse transcriptase was added, mixed by pipetting and incubated at 42°C for 90 minutes followed by 70°C for 10 minutes.

To 1 µl aliquot of the first strand cDNA, 10 µl of 5X Primestar buffer, 4 µl of phosphorylated 10 µM PCR primer M1, 4 µl of dNTP mix of 2.5 mM each and 0.5 µl of 2.5 U Primestar HS DNA polymerase was added, made upto 50 µl with sterile water and subjected to following PCR cycles: initial denaturation of 95°C for 1 minute followed by 18 cycles of 95°C for 5 seconds, 65°C for 5 seconds and 68°C for 6 minutes. The phosphorylated PCR primer M1 was prepared by mixing 1 µl of 100 µM PCR primer M1, 1 µl of 10X T4 DNA ligase buffer, 0.5 µl of 10 U/µl T4 polynucleotide kinase and 7.5 µl of sterile water followed by incubation at 37°C for 60 minutes and heat inactivation by incubating at 65°C for 25 minutes. After the PCR reaction, the ds cDNA was purified using PCR purification as mentioned in section 3.8.3.

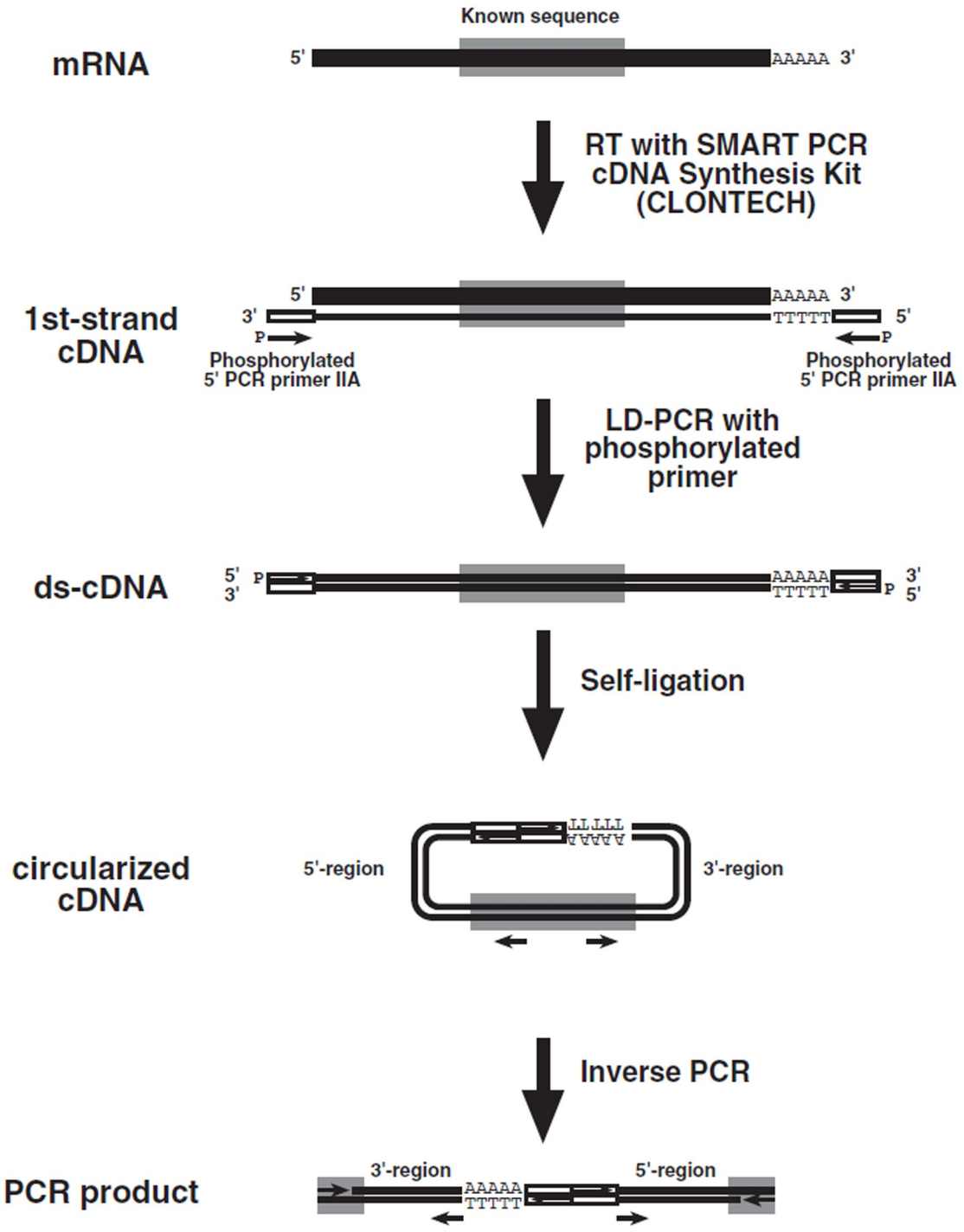


Figure 9. Schematic outline of SMARTer ds cDNA-inverse PCR approach for full length cDNA amplification (Kuniyoshi et al., 2006)

Inverse PCR was performed with an aliquot of 50 ng ds cDNA mixed with 5 µl of 10X T4 DNA ligase buffer and 1 µl of 3U T4 DNA ligase and made upto 10 µl with sterile water, followed by incubation at 22°C for 30 minutes and overnight incubation at 4°C for self circularization of ds cDNA. Two microlitre of self circularized cDNA, 5 µl of 10X ExTaq buffer, 4 µl of dNTP mix of 2.5 mM each, 10 µM outward primer 1, 10 µM outward primer 2 and 0.25 µl of 5 U ExTaq hotstart DNA polymerase was made upto 50 µl with sterile water and subjected to following PCR cycles: initial denaturation at 95°C for 1 minute, 30 cycles at 95°C for 30 seconds, 62°C for 30 seconds and 72°C for 2 minutes and final extension at 72ºC for 10 minutes. A negative control was kept with non circularized ds cDNA as the template. A nested PCR was also done with the same PCR parameters with 2 µl of first inverse PCR product and with a set of nested outward primers. The amplified PCR products were mixed with 5µl of 6X gel loading dye and resolved in 1% agarose gel and gel documented. The DNA band unique to circularized cDNA template with high molecular weight was excised and DNA was extracted from the gel by gel extraction protocol.

3.13.2 Extraction of DNA from agarose gel

The amplicon of expected size was sliced from 1% agarose gel by viewing under UV transilluminator and DNA was extracted from the gel using Genelute gel extraction kit (Sigma). Three volumes of gel solubilization solution was added to the sliced gel and incubated at 50°C for 5-10 minutes till the gel dissolved completely in the buffer. After the gel dissolved completely, one volume of isopropanol was added and mixed by inversion. The spin column was placed in 2 ml collection tube, 500 µl of column preparation solution was added and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and the solubilized gel solution was added to the spin column in volumes of 700 µl and repeatedly centrifuged at 13000 rpm for 1 minute till all the solubilized gel solution passed through the spin column. The spin column was washed with 600 µl of wash buffer and centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and dry spin was done by centrifuging the spin column at 13000 rpm for 2 minutes. The spin column was placed in a new 1.5 ml centrifuge tube and 30 µl of sterile water was added to the centre of the column membrane. The spin column was allowed to stand at room temperature for 1 minute and centrifuged at 13000 rpm for 1 minute to elute the DNA fragments from the spin column.

3.13.3 DNA sequencing and sequence analysis

The eluted DNA fragments were ligated into pGEM-T Easy vector (Promega) as mentioned in section 3.8.6. The volume of eluted DNA taken for ligation was calculated using the formula: (Concentration of vector in ng × size of insert in kb) / (size of vector in kb × insert:vector molar ratio), where 50 ng of 3.0 kb vector was used throughout the experiment with insert:vector molar ratio of 3:1. The ligated product was transformed into *E. coli* JM109 as mentioned in section 3.8.6. The recombinants were confirmed; plasmids were isolated and sequenced as mentioned in section 3.8.7. The sequence analysis was done with Bioedit software (Hall, 1999). The sequences obtained with both forward and reverse primers were aligned using 'Cap contig assembly' in Bioedit and the vector sequences were removed using Vecscreen. The trimmed sequence was again trimmed at adaptor position (PCR primer M1) and aligned each fragment with the template sequence used for inverse PCR primer designing. The 'Cap contig assembly' was done to align the sequences. An end to end PCR was also carried out to confirm the sequence with primers designed from the ends of the aligned sequence.

3.14 Sequence analysis of full length cDNA

The full length cDNA sequence was translated using the EMBOSS TRANSEQ program (Rice et al., 2000) to obtain deduced amino acid sequences. UTRscan (http://itbtools.ba.itb.cnr.it/utrscan) and RegRNA analysis (http://regrna.mbc.nctu.edu.tw/html/references.html) were done to identify the UTR motifs. Comparative sequence analysis was performed using blastp (Altschul et al., 1990). The open reading frame (ORF) was predicted by ORF Finder (Rombel et al., 2002). The physical and chemical parameters of the protein sequences were analysed using ProtParam tool (Gasteiger et al., 2005). Subcellular localisation analysis and the presence of signal peptide and transmembrane regions were analysed using SLP-local (Matsuda et al., 2005), SignalP 4.1 server (Petersen et al., 2011) and TMPred server (Hofmann and Stoffel, 1993) respectively. Multiple sequence alignment of protein sequences with other orthologous sequences was conducted with ClustalW (Thompson et al., 1994) using default parameters. Phylogenetic analysis of sequences was performed using different methods such as Bayesian analysis, maximum parsimony and maximum likelihood in Phylip package and the consensus tree was generated. The amino acid residues of substrate binding pocket, cyclization pocket and geometry shapers were searched using the SBSPKS web server (Anand et al., 2010).

3.15 Molecular modelling and docking

Three-dimensional (3D) protein structure was constructed using Modeller 9.10 package (Eswar et al*.,* 2006). Calculation of cavity volumes was performed with the CASTP program (Dundas et al., 2006) with default probe radius of 1.4 Angstroms and the 3D molecular structures and active-sites of proteins were visualized with UCSF Chimera (Huang et al., 1996). The side chains and hydrogen atoms added for refining the structure and the stability of homology model has been validated by checking the geometry using PROCHECK (Laskowski et al., 1993). The model was validated using Ramachandran plot. Ramachandran plot was identified by Procheck program of Structural Analysis and Verification Server (Laskowski et al., 1993). Molecular docking study was carried out using Molegro Virtual Docker (Thomsen and Christensen, 2006) with the canonical smile notations of substrates collected from PubChem (Kim et al., 2015). The 3D structures of compounds were developed by 3D Structure Generator CORINA (Sadowski et al., 1994) using canonical smiles of the compound.

3.16 Pre-mRNA analysis

To detect the possibility of post-transcriptional silencing, three primers were designed; forward primer from exon 1 region, one reverse primer from intron region and another reverse primer from exon 2 region. The primer combination exon1-intron was used to analyze the pre-mRNA and exon 1-exon 2 primers were used for mature mRNA detection with first strand cDNA as the template. Primer designing and PCR reaction was done as described in section 3.6. A reverse transcription negative control (without reverse transcriptase) and a non-template negative control were included to confirm the absence of genomic DNA and to check non-specific amplification respectively.

3.17 Analysis of upstream sequence of coding region

The genomic DNA was isolated according to Syamkumar et al. (2003). Five grams of leaves were ground using mortar and pestle with liquid nitrogen and a pinch of PVP. Twenty millilitre of extraction buffer (100 mM Tris-pH 8, 20 mM EDTA-pH 8, 1.4 M NaCl and 2% CTAB) pre-warmed at 65°C for 20 minutes was added to the ground tissue, transferred to 50 ml centrifuge tube and incubated at 65°C for 30 minutes. After the incubation, an equal volume of chloroform:isoamylalcohol (24:1) was added and centrifuged at 3000 rpm for 20 minutes. The aqueous phase was transferred to fresh tube

and 0.6 volumes of isopropanol was added and incubated overnight at -20°C and centrifuged at 3000 rpm for 6 minutes. The pellet was air dried and dissolved in 500 µl of sterile water. To this, 5 µl of 10 mg/ml RNase was added and incubated at 37°C for 30 minutes followed by proteinase treatment with 5 µl of 1 mg/ml proteinase K and incubated at 37°C for 30 minutes. After the incubation, equal volume of phenol: chloroform: isoamylalcohol (25:24:1) was added and mixed by inversion and centrifuged at 14000 rpm for 10 minutes. Equal volume of chloroform: isoamylalcohol (24:1) was added to the aqueous phase and centrifuged at 14000 rpm for 10 minutes. To the aqueous phase, 2.5 volumes of ice cold ethanol was added and incubated overnight at -20°C. The DNA-ethanol mix was centrifuged at 14000 rpm for 10 minutes and 500 µl of 70% ethanol was added to the pellet, mixed by pipetting and centrifuged at 10000 rpm for 5 minutes. The pellet was air dried and dissolved in 200 µl of sterile water.

Purified genomic DNA (500 ng) was mixed with 1 µl of 10 U/µl *Pst*I and 2 µl of 10X buffer O and made upto 20 µl with sterile water and incubated at 37°C for 10 minutes. After the incubation, the restricted products were purified using PCR purification kit as described in section 3.8.3. The purified restricted DNA (10 ng) was mixed with 1 µl of 3 U/µl T4 DNA ligase and 2 µl of 2X T4 DNA ligase buffer and incubated at 22°C for 30 minutes followed by incubation at 4°C overnight. The inverse PCR and sequence analysis and was performed as described in section 3.13. The transcriptional start site, *cis*-regulatory elements and MYB binding sites in the sequences upstream of coding sequence was predicted using online server TSSPlant (http://www.softberry.com/berry.phtml?topic=tssplant&group=programs&subgroup=promoter), PlantCARE database (Lescot et al., 2002) and PlantRegMap (Jin et al., 2017) respectively.

3.18 Isolation of protein from turmeric rhizomes

The turmeric rhizomes harvested at 120 DAP were used for protein isolation following the protocol of Amalraj et al. (2010). The rhizomes were washed thoroughly and ground immediately using mortar and pestle with liquid nitrogen. One millilitre of extraction buffer (0.1 M Tris-pH 8, 10 mM EDTA-pH 8, 0.9 M sucrose and 0.1% DTT) was added to 100 mg of the powdered sample taken in a 2 ml centrifuge tube. Equal volume of Tris saturated phenol was added, vortexed for 30 seconds and incubated at 28ºC for 30 minutes in a rotary shaker with 150 rpm. After the incubation, the sample was centrifuged at 13000 rpm for 20 minutes at 4ºC. The upper phase of 300 µl each was

transferred to 1.5 ml centrifuge tube and 4 volumes of methanol containing 0.1 M ammonium acetate were added and the samples were incubated overnight at -20ºC. The samples were then centrifuged at 13000 rpm for 20 minutes at 4ºC. One millilitre of methanol containing 0.1 M ammonium acetate was added, incubated in -20ºC for 30 minutes and centrifuged at 13000 rpm for 10 minutes at 4ºC. The supernatant was discarded and the pellet was resuspended in 80% acetone, incubated at -20ºC for 30 minutes and centrifuged at 13000 rpm for 10 minutes at 4ºC. This step was repeated with 80% acetone and then with 70% ethanol. The pellet was air dried at room temperature for 15 minutes and resuspended in 50 µl of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS and 30 mM Tris). The solubilised pellet was again centrifuged and the clear supernatant was transferred to a new 0.5 ml centrifuge tube and purified using 2D clean up kit (GE healthcare) prior to 2D electrophoresis. Briefly, 150 µg of protein was mixed with 300 µl of precipitant and incubated on ice for 15 minutes. After the incubation, 300 µl of co-precipitant was added and centrifuged at 13000 rpm for 15 minutes. The supernatant was discarded and 40 µl of co-precipitant was added to the pellet, incubated on ice for 5 minutes and centrifuged at 13000 rpm for 5 minutes. Twenty five microlitre of sterile water was added to the pellet and vortexed for 5-10 seconds. One millilitre of wash buffer pre-chilled at -20ºC and 5 µl of wash additive was added, incubated at -20ºC for 1 hour with vortexing for 20 seconds every 10 minutes and centrifuged at 13000 rpm for 5 minutes at 4ºC. The pellet was air dried for 5 minutes and re-suspended in 100 µl of lysis buffer (8 M urea, 2% CHAPS, 0.5% IPG buffer, 40 mM DTT and 0.02% bromophenol blue).

3.18.1 Determination of concentration and purity of proteins

Briefly, 10 µl of protein sample, 1990 µl of sterile water and 2 ml of concentrated Bradford reagent (Sigma) were mixed and the absorbance of the solution was read at 595 nm using UV/VIS spectrophotometer. The blank was prepared with 2 ml of concentrated Bradford reagent and 1 ml of sterile water. Bovine Serum Albumin (Sigma) was used as the standard. An increasing concentration of 5-30 µg of Bovine Serum Albumin was mixed with 2 ml of concentrated Bradford reagent and made upto 3 ml with sterile water. The standard curve was prepared by plotting the absorbance of the standards read at 595 nm against the concentration of the protein. The standard curve was used to determine the concentration of protein solution.

The quality of the protein was assessed by SDS PAGE (Laemmli, 1970) using Protean II Xi (Bio-Rad) vertical electrophoresis apparatus. The spacers (0.5 mm), outer and inner glass plates were washed with mild soap solution and tap water, rinsed two times with distilled water and wiped with lint free tissue paper. The spacers were placed in between the outer and inner plate and screwed using sandwich clamps and the whole assembly was placed into the casting stand. Resolving gel (12%) was prepared by mixing 21.45 ml of sterile water, 15 ml of 40% acrylamide-bisacrylamide mix, 12.5 ml of 1.5 M Tris HCl-pH 8, 500 µl of 10% SDS, 50 µl of TEMED and 500 µl of freshly prepared 10% APS. The monomer solution was immediately overlaid with isopropanol and allowed to polymerize for 30 minutes. Resolving gel was poured such that it covers $3/4^{th}$ of the resulting gel and was allowed to polymerize for 30 minutes. After the gel gets solidified, the overlaid solution was poured off and rinsed several times with distilled water. The stacking gel was poured onto the polymerized gel. The stacking gel (4%) was prepared by mixing 7.8 ml of sterile water, 1.25 ml of 40% acrylamide-bisacrylamide mix, 3.15 ml of 0.5 M Tris HCl-pH 6.8, 125 µl of 10% SDS, 12.5 µl of TEMED and 125 µl of freshly prepared 10% APS. The comb was immediately placed into the gel sandwich and allowed to polymerize for another 30 minutes. After the gel gets polymerized, the comb was carefully removed and the wells were washed with sterile water. The assembled gel unit was then placed in electrophoresis tank filled with running buffer (192 mM Tris, 25 mM glycine and 0.1% SDS). The protein sample (50 µg) to be loaded into the well was made upto 15 µl with sterile water and equal volume of 2X sample loading dye (0.125 M Tris-pH 6.8, 4% SDS, 5% β-mercaptoethanol, 20% glycerol and 0.01% bromophenol blue) was added and denatured by incubating at 85ºC for 3 minutes. After the incubation, the samples were snap cooled on ice and loaded into the wells and electrophoretially separated at 8mA till the bromophenol blue dye front reached the end of the gel. Protein molecular weight marker (Takara) was also loaded in a well for comparison of molecular mass of peptides.

3.19 Two dimensional (2D) electrophoresis

2D electrophoresis was done using pre-cast IPG gel strips of pH 4-7, 13 cm (GE healthcare) on Ettan IPGphor unit (GE healthcare). One hundred and fifty mirogram of protein was dissolved in rehydration buffer (8 M urea, 2% CHAPS, 0.5% IPG buffer, 40 mM DTT and 0.02% bromophenol blue) to a final volume of 340 µl and was applied to the cleaned immobiline strip tray. The plastic cover on the strip was removed carefully with

the forceps and the strip was placed over the sample in the immobiline strip tray without the formation of any air bubble ensuring that entire length of the strip is in touch with the sample and the gel side of the strip faces the sample. The sample was overlaid with dry strip cover fluid and allowed to passively rehydrate at room temperature for 16 hours.

The IPGphor bed was cleaned and the level of the manifold was checked using round spirit level. Dry strip cover fluid (108 ml) was added to the manifold placed on the IPGphor bed. The rehydrated strip was transferred to the manifold ensuring that the gel side faces up, the + end of the strip placed in the top on the IPGphor bed and 13 cm gel touched the appropriate etched portion. One hundred and fifty microlitre of sterile water was added to paper wicks and one each was kept at both the ends of the strip ensuring that it touched the gel. The electrodes were placed over the paper wick which is in contact with the strip gel. The electrodes were tightened, the lid was closed and electrofocused for 36 kvh at 20ºC with 50 µA/strip. The parameters used were step 1- step and hold at 500 V for 500 Vhr, step 2- gradient at 1000 V for 800 Vhr, step 3- gradient for 10000 V for 16500 Vhr and step 4- step and hold at 10000 V for 13700 Vhr.

After electrofocussing, the strips were equilibrated with 15 ml of reduction buffer (6 M urea, 2% SDS, 0.002% bromophenol blue, 30% glycerol, 50 mM Tris HCl-pH 8.8 and 1% DTT) for 30 minutes with gentle shaking at 37ºC. The strips were then transferred to 15 ml of alkylation buffer (6 M urea, 2% SDS, 0.002% bromophenol blue, 30% glycerol, 50 mM Tris HCl-pH 8.8 and 2.5% iodoacetamide) for 30 minutes with gentle shaking at 37ºC. The strips were then rinsed with running buffer (192 mM Tris, 25 mM glycine and 0.1% SDS). Acrylamide gel (12%) was prepared as described in section 3.18.1 and was poured into the inner plate leaving 1 cm for placing the strip. After the gel gets solidified, running buffer was poured onto the top of the gel and the equilibrated strips were carefully placed on the gel such that no air bubble is trapped inside. The running buffer was carefully removed. One percentage of agarose in running buffer containing a pinch of bromophenol blue was melted, cooled to 50ºC and poured onto the strip placed on the gel. The assembled gel unit was then placed in electrophoresis tank filled with running buffer (192 mM Tris, 25 mM glycine and 0.1% SDS) and electrophoretially separated at 8 mA till the bromophenol blue dye front reached the end of the gel.

3.19.1 Gel staining and image acquisition

After the electrophoresis, the gel was carefully detached from the glass plates by placing gel with glass plates in a tray with distilled water and detached using spacers. The gel was then fixed for 1 hour with solution 1 (40% ethanol, 10% glacial acetic acid and 10% methanol). The gel was then sensitized for 2 hours with solution 2 (1% acetic acid and 10% ammonium sulfate). After the treatment, the gel was stained overnight with solution 3 (5% acetic acid, 45% ethanol and 0.125% coomassie brilliant blue R250) on a rotary shaker. After the staining, the gel was treated with solution 4 (5% acetic acid and 40% ethanol) for 1 hour and in solution 5 (3% acetic acid and 30% ethanol) till the background stain was removed from the gel. The gel was then used for image acquisition and further stored in 5% acetic acid.

Results...

Results

4.1 RNA isolation from different tissues of *Curcuma longa*

Isolation of high quality RNA was an essential prerequisite due to the reason that turmeric is loaded with extractable polyphenols and polysaccharides and a universal protocol which was rapid and economical for downstream analysis like NGS, qRT-PCR etc. was needed. So four protocols already available for isolation of RNA from recalcitrant tissues rich in polyphenols and polysaccharides (Chomczynski and Sacchi, 1987; Hou et al., 2011; Chang et al., 1993 and Ghawana et al., 2011) were attempted to optimize a suitable one for turmeric. Good quality RNA was isolated from leaves with highest yield of RNA (164±7.3 µg/g) obtained using Trizol method (Chomczynski and Sacchi, 1987) and least (17±3.5 µg/g) recovered following Ghawana et al. (2011). However, in case of rhizomes, no RNA was obtained by following Chomczynski and Sacchi (1987) and Hou et al. (2011). Since rhizomes are rich in polysaccharides, as recommended in trouble shooting guidelines of Trizol method for tissues rich in polysaccharides, the aqueous phase obtained after centrifuging the homogenized tissue with Trizol reagent was precipitated using 0.25 volumes of 1.2 M sodium citrate, 0.8 M NaCl and 0.25 volumes of isopropanol. However, this procedure failed to extract RNA from rhizomes. Moderate yield of RNA of 31±6.2 µg/g and 29±2.7 µg/g was observed using protocol by Ghawana et al. (2011) and Chang et al. (1993) respectively. Plant RNeasy kit (Qiagen) was also used to extract RNA from rhizomes, but the yield was too low (6.1±0.06 µg/g). The results are presented in **Table 1** and **Fig. 10**. DNA contamination was present in RNA isolated using all the tested protocols with the least observed in the protocol of Chang et al. (1993), which involves the addition of lithium chloride and overnight incubation. But this approach is time consuming (~16 h). In this backdrop, developing a quick, handy and cheap protocol for isolating RNA from different tissues of turmeric with good yield and quality was realised to be vital.

4.1.1 Modified protocol for total RNA extraction

The protocol suggested by Hou et al. (2011) was selected for modification as it is simple, rapid and even a change in the sample: extraction buffer volume from 1:1.2 to 1:20 yielded RNA of comparable quantity and quality (24.4 µg/g; A_{260}/A_{280}:1.71).

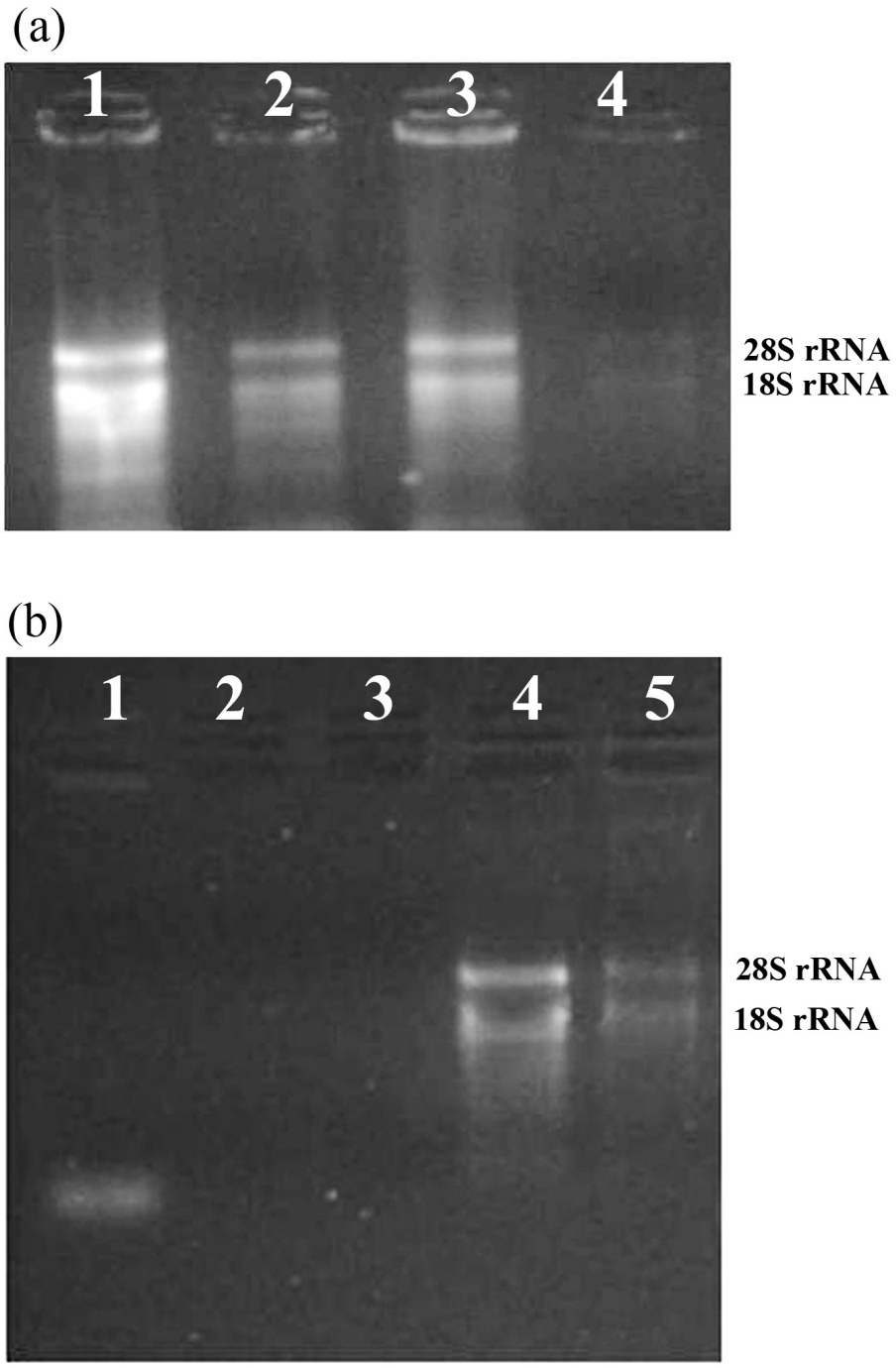


Figure 10. Agarose gel electrophoresis of RNA (5 µl each sample) isolated from
(a) leaves using using different protocols, where lane (1) Chomczynski and Sacchi, 1987, (2) Hou et al., 2011, (3) Chang et al., 1993, (4) Ghawana et al., 2011 and
(b) rhizomes using different protocols, where lane (1) Chomczynski and Sacchi, 1987, (2) Plant RNeasy kit, (3) Hou et al., 2011, (4) Chang et al., 1993, (5) Ghawana et al., 2011.

The extraction buffer mentioned in Hou et al. (2011) contains Tris-HCl, sodium chloride, SDS and β-mercaptoethanol. But the presence of sodium chloride in the extraction buffer decreased the quality of RNA.

Table 1. Spectrophotometric analysis of yield and purity of RNA isolated from rhizomes and leaves using different protocols

Method	**Tissue used**	**$A_{260/280}$**	**RNA yield (µg/g)**
Chomczynski and Sacchi (1987)	Leaf	1.73±0.80	164±7.3
Hou et al. (2011)	Leaf	1.60±0.30	40±2.5
Chang et al. (1993)	Leaf	1.99±0.02	27±2.0
Ghawana et al. (2011)	Leaf	2.20±0.44	17±3.5
Chomczynski and Sacchi (1987)	Rhizome	-	-
Hou et al. (2011)	Rhizome	-	-
Chang et al. (1993)	Rhizome	1.92±0.05	29±2.7
Ghawana et al. (2011)	Rhizome	1.63±0.07	31±6.2
Plant RNeasy kit	Rhizome	1.50±0.07	6.1±0.06

The concentration of SDS (0–3%) and β-mercaptoethanol (0–3%) was optimized individually in the protocol. The yield of RNA was higher when 2% SDS (137.6 µg/g) was used compared to the rest. The concentration of 1% β-mercaptoethanol was found to be optimum compared to the rest. The addition of PVP to the extraction buffer also increased the RNA yield and its concentration was optimized. The yield of RNA was the best when 2% PVP (259.4 µg/g) was used in comparison with the rest. For the enhancement of RNA precipitation, sodium acetate was added along with isopropanol. Re-extraction of the aqueous phase along with the sodium acetate and acid phenol:chloroform decreased the polysaccharide and protein contamination and increased the RNA yield. When the concentration of sodium acetate (1 M-5 M) was increased gradually, yield of RNA also increased. Further to enhance the RNA precipitation and for removal of polysaccharides, isopropanol was added along with sodium acetate (0-5 M) and 3 M sodium acetate was found to be optimum (59.16 µg/g) (**Table 2**, **Fig. 11**). The effect of pH of extraction buffer was also monitored. The yield was slightly higher in acidic pH (272.4 µg/g at pH 5) when compared with basic pH (229.2 µg/g at pH 8). The RNA yield was higher with 1 hour incubation after the addition of isopropanol to the aqueous phase (228.66 µg/g) when compared with 10 minutes incubation (33.5 µg/g).

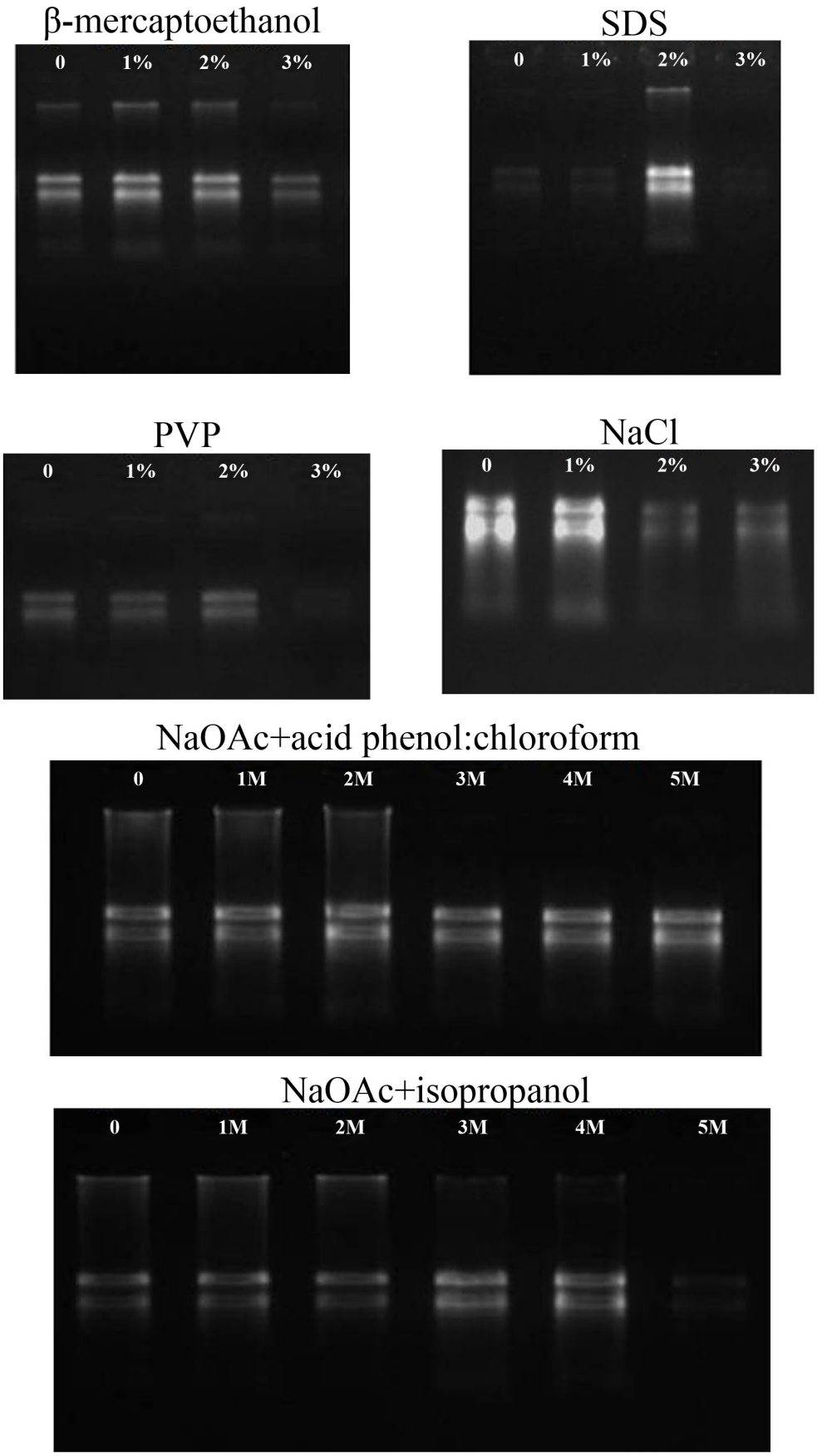


Figure 11. Optimization of major components in RNA isolation

Table 2. Evaluation of yield and quality with different concentration of components in RNA isolation

	Concentration	$A_{260/280}$	RNA yield (µg/g)
NaCl	-	1.69	154.3
	1 M	1.61	133.2
	2 M	1.59	121.6
	3 M	1.52	116.7
SDS	-	1.59	89.7
	1 %	1.63	80.6
	2 %	1.62	137.6
	3 %	1.55	107.4
β-mercaptoethanol	-	1.59	84.3
	1 %	1.64	120.0
	2 %	1.62	73.9
	3 %	1.62	89.5
PVP	-	1.79	159.0
	1 %	1.83	228.8
	2 %	1.76	259.4
	3 %	1.73	105.4
Sodium acetate (along with acid phenol: chloroform)	-	1.24	93.2
	1 M	1.49	118.2
	2 M	1.82	104.4
	3 M	1.87	106.8
	4 M	1.87	58.1
	5 M	1.75	110.2
Sodium acetate (along with isopropanol)	-	1.54	15.7
	1 M	1.76	41.4
	2 M	1.79	53.6
	3 M	2.14	59.2
	4 M	1.98	62.2
	5 M	2.02	66.0

The optimised protocol is as below:

Grind 100 mg of sample with liquid nitrogen

↓

Add PVP (2% of extraction buffer) while grinding

↓

Add 2 ml of extraction buffer pre-warmed at 65 °C

(2% SDS, 100 mM Tris-HCl - pH 8, 25 mM EDTA- pH 8 and 1% β-mercaptoethanol)

↓

Transfer the contents equally to two, 2 ml centrifuge tubes

↓

Add equal volume of acid phenol: chloroform and vortex for 10 s

↓

Incubate at room temperature for 10 min

↓

Centrifuge at 15000 g for 10 min at 4°C

↓

Transfer the supernatant to fresh tubes

↓

Add 0.3 volumes of 5 M sodium acetate and 0.7 volumes of acid phenol:chloroform

↓

Mix by inverting the tubes and incubate in ice for 10 min

↓

Centrifuge at 15000 g for 10 min at 4°C

↓

Transfer the supernatant to fresh 1.5 ml centrifuge tubes

↓

Add 0.1 volume of 3 M sodium acetate and equal volume of isopropanol

↓

Incubate the tubes at -20 °C for 1 h

↓

Centrifuge at 15000 g for 10 min at 4°C and discard the supernatant

↓

Wash pellet with 70% ethanol by centrifuging at 7500 g for 5 min at 4°C

↓

Air dry the pellet for 5-10 min and dissolve in 25 µl RNase free water

4.1.2 Determination of RNA concentration and purity

The optimized protocol for RNA isolation worked well with all the tissues of turmeric namely; leaves, pseudostem, rhizomes and roots and the concentration of RNA was analyzed using Biophotometer plus (Eppendorf). The highest concentration of RNA was extracted from rhizomes (180±13µg/g), followed by leaves, shoots and roots. A_{260}/A_{280} ratio ranged from 1.65 to 1.89 and A_{260}/A_{230} ratio exceeded 1.77 (**Table 3**). The integrity of RNA was examined on 1.2% denaturing TAE/agarose gel electrophoresis and it showed sharp, intact 28S and 18S rRNA bands (**Fig. 12**). Additional rRNA bands were

observed in leaves and pseudostem tissues. RIN values of the RNA isolated were found to be 7.1, 7.9, 7.8 and 8.2 from leaves, pseudostem, rhizomes and roots respectively (**Fig. 13**).

4.1.3 Isolation of low molecular weight RNA from fresh rhizomes of turmeric, total RNA from rhizomes of other members of Zingiberaceae and tissues stored in RNAlater

To analyze whether the optimized RNA extraction protocol retains small RNAs, the small RNA fraction was enriched from total RNA of leaf, pseudostem, rhizome and root tissues using polyethylene glycol and sodium chloride. Intact bands corresponding to 5.8S rRNA, 5S rRNA, tRNA and microRNAs were observed when 10 µg of RNA was loaded on 15% denaturing PAGE (**Fig. 14**). The optimized total RNA isolation protocol worked well in three genotypes of *C. longa* (Collection no. 19, RH17 and Megha turmeric), three different species of *Curcuma* (*C. xanthorrhiza*, *C. amada* and *C. caesia*) and ginger (*Zingiber officinale*). Intact 28S and 18S rRNA bands were observed in 1.2% agarose gel. Highest yield of RNA (256.45 µg/g) was obtained for *C. xanthorrhiza* and least (132.10 µg/g) for Megha turmeric (**Table 3**, **Fig. 15**).

Table 3. Quality and yield of RNA extracted using optimized protocol from different tissues of turmeric and related spp.

	Tissue used	**$A_{260/280}$**	**RNA yield (µg/g)**
C. longa (IISR Prathibha)	Rhizome	1.80 ± 0.05	180 ± 13
	Leaf	1.82 ± 0.02	163 ± 50
	Shoot	1.87 ± 0.02	138 ± 6.5
	Root	1.75 ± 0.10	102 ± 7.2
C. longa (Collection no. 19)	Rhizome	1.81	133.30
*C. longa (*RH17)	Rhizome	1.79	142.60
C. longa (Megha turmeric)	Rhizome	1.82	132.10
C. xanthorrhiza	Rhizome	1.85	256.45
C. amada	Rhizome	1.81	134.00
C. caesia	Rhizome	1.75	146.85
Zingiber officinale	Rhizome	1.87	215.65

The RNA yield was lesser from tissues stored in RNAlater and also a slimy precipitate was observed in the final step. To avoid the slimy pellet in the final step, optimized protocol for isolating RNA was followed and the supernatant obtained after centrifuging with 0.3 volumes of 5 M sodium acetate and 0.7 volumes of acid phenol: chloroform was transferred to a spin column, centrifuged at 10000 g for 1 minute at 4ºC and the eluate was discarded.

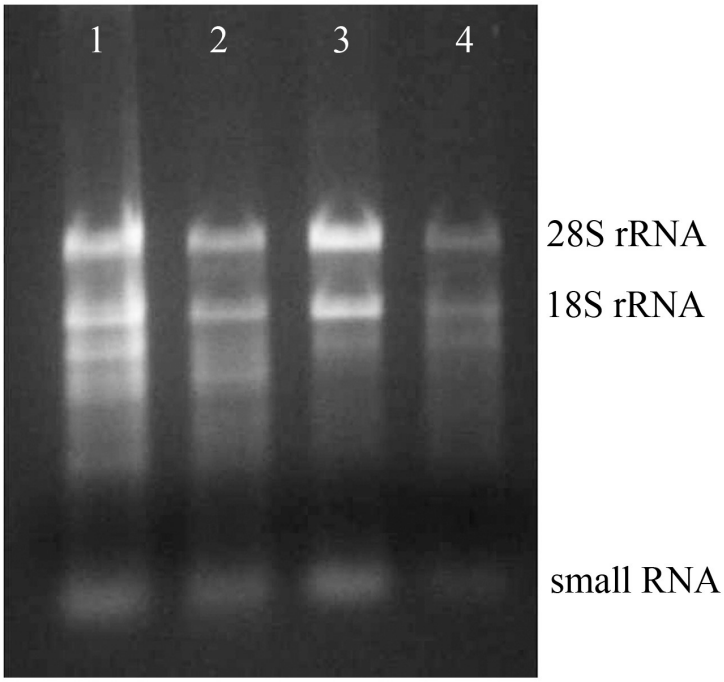


Figure 12. Denaturing agarose gel electrophoresis of RNA isolated from leaves (1), pseudostem (2), rhizomes (3) and roots (4)

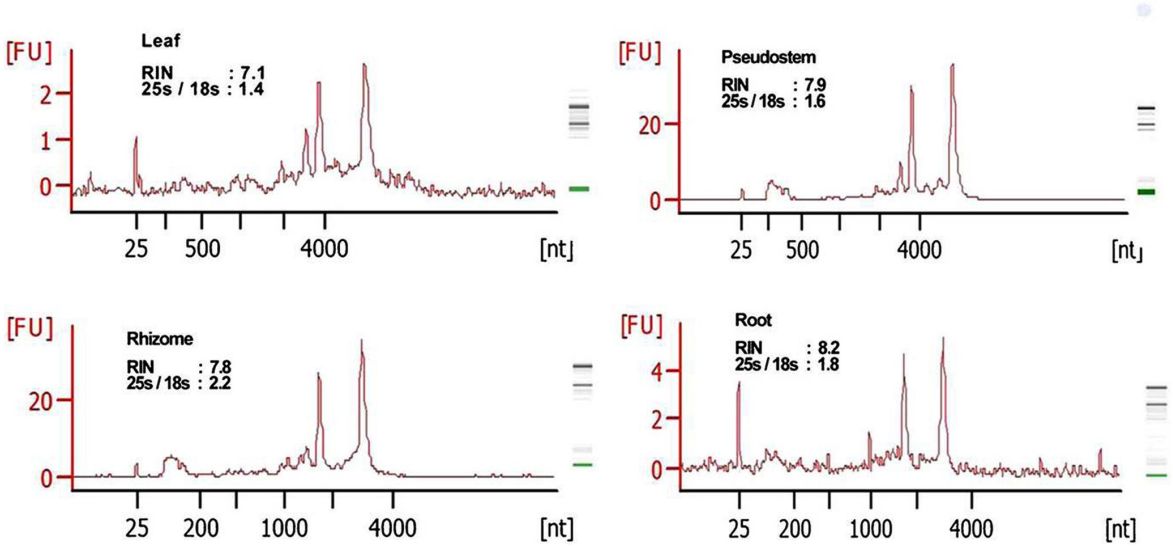


Figure 13. Electropherogram of total RNA isolated from different tissues of turmeric generated by Agilent 2100 Bioanalyzer

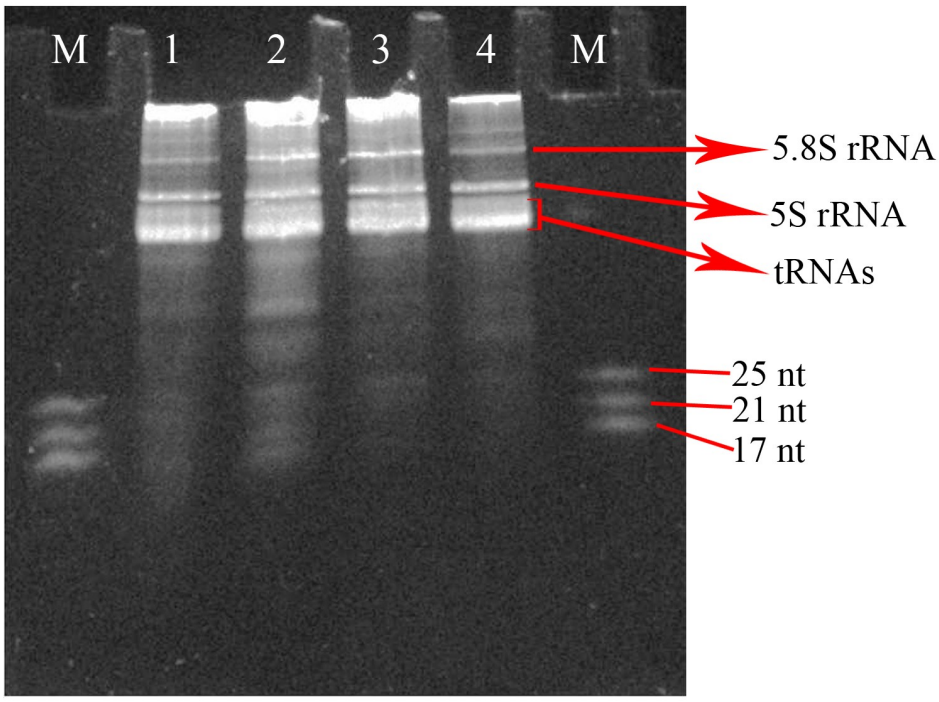


Figure 14. Denaturing polyacrylamide gel electrophoresis (15%) of small RNA isolated from leaves (1), pseudostem (2), rhizome (3) and root (4), Lane M- microRNA marker

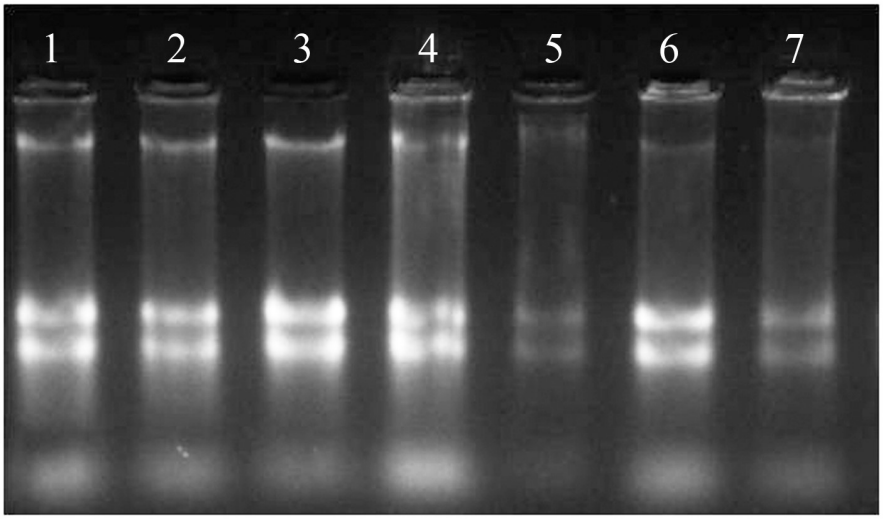


Figure 15. Agarose gel electrophoresis of RNA isolated from rhizomes of *C. longa* Collection no. 19 (1), *C. longa* RH 17 (2), *C. longa* Megha turmeric (3), *C. xanthorrhiza* (4), *C. amada* (5), *C. caesia* (6) and *Zingiber officinale* (7)

This step was repeated till the complete supernatant passed through the spin column. The spin column was then washed twice with wash solution (1:5 dilution of 50 mM Tris HCl-pH 8 with 100% ethanol) following the same centrifugation conditions. The spin column was then centrifuged at 10000 g for 2 minutes at 4ºC to remove the remaining wash solution. The spin column was transferred to new 1.5 ml centrifuge tube. Nuclease free water (30 µl) was added to the centre of the column and incubated at 65ºC for 2 minutes and centrifuged at 10000 g for 2 minutes at 4ºC. The yield of RNA from leaves and rhizomes were 35 µg/g and 109 µg/g respectively.

4.2 cDNA synthesis and qRT-PCR

The efficiency of RNA for application in several downstream processes like cDNA synthesis and qRT-PCR was assessed. cDNA was synthesised from RNA extracted from rhizomes using the new protocol as well as other protocols (Ghawana et al., 2011 and Chang et al., 1993). RNA of leaves, pseudostem and roots extracted using our protocol was also reverse transcribed. As shown in **Fig. 16**, ~950 bp of *ACTIN* was successfully amplified from all these cDNA samples using *ACTIN* primers (Appendix 1). The cDNAs synthesized from RNA extracted from rhizomes by all the three protocols were compared using qRT-PCR with *ACTIN* primers. All the samples showed amplification with normal Cq values with the standard deviations in the range that allowed for meaningful analysis of expression data and a single peak was observed in the melting curve (**Fig. 17**). Reverse transcription negative control was also analyzed for each sample and expression was not observed using cDNA synthesized from RNA using optimized protocol till 35 cycles. In order to confirm the extraction of small RNAs, qRT-PCR was done with primers specific to a novel turmeric rhizome specific microRNA (Santhi and Sheeja, 2013) and cDNAs synthesized from RNA extracted from leaves, pseudostem, rhizomes and roots as the template. The rhizome and root tissues showed amplification but leaf and pseudostem tissues failed to show any expression (**Fig. 18**).

4.3 Construction of full length normalized cDNA library

Turmeric is a non model crop and genomic resources available are limited to ESTs. To provide a resource of four month old turmeric specific expressed sequence data and to mine the full length candidate genes for curcumin biosynthesis, a full length normalized cDNA library was synthesized. Leaf, pseudostem, rhizome and root tissues of four month

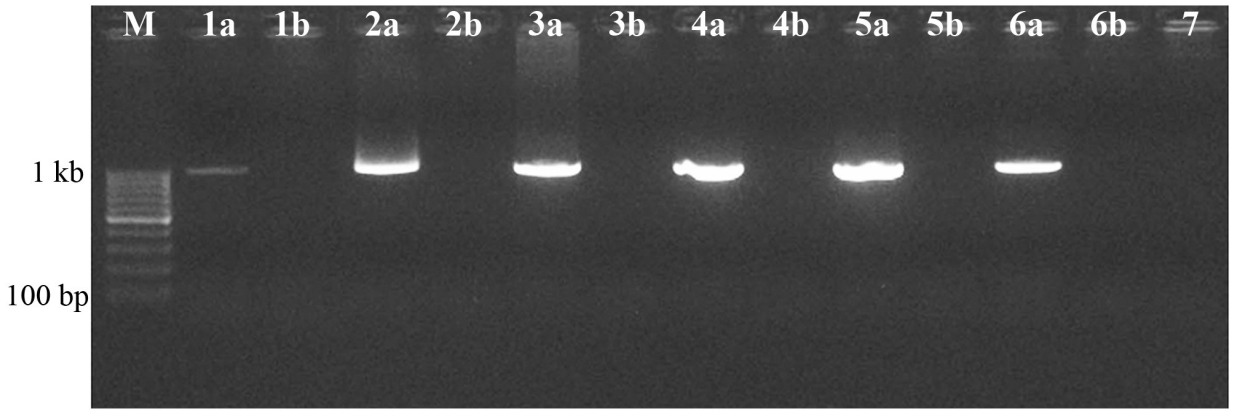


Figure 16. PCR amplification of actin using the cDNA from rhizome following Ghawana et al., 2011 (1a), Chang et al. 1993 (2a) and the new protocol (5a). Lanes 3a, 4a and 6a correspond to actin amplification from cDNA of leaf, pseudostem and root respectively following the new protocol. 1b - 6b correspond to the reverse transcription negative control of the samples 1a - 6a respectively, 7- non template control. M-100 bp DNA ladder

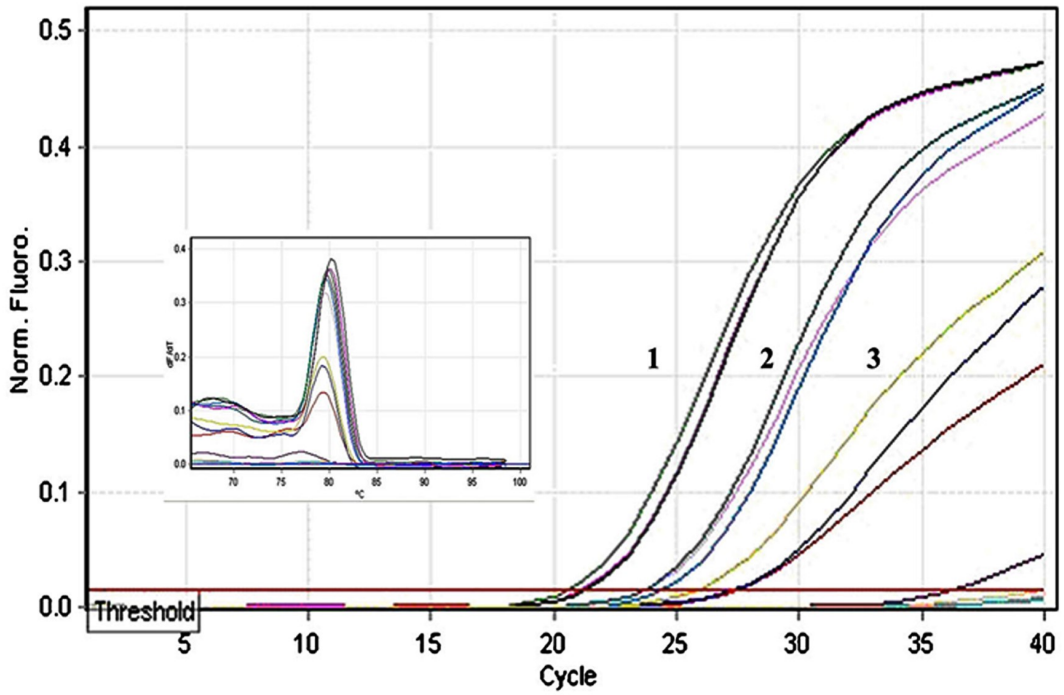


Figure 17. Quality check of cDNA by qRT-PCR generated from rhizome using 1- New protocol (Mean Ct- 22), 2- Chang et al. 1993 (Mean Ct- 24.6), 3- Ghawana et al. 2011 (Mean Ct- 28.5)

old high curcumin turmeric variety, Megha turmeric were pooled for construction of first strand cDNA using SMARTer (Switching Mechanism At 5′-end of RNA Transcript) PCR cDNA synthesis kit. ds cDNA was synthesized from first strand cDNA using advantage 2 polymerase mix and PCR primer M1 (Appendix 1) by subjecting to 15-18 cycles of PCR. When the PCR products were electrophorized, a smear like pattern was observed in all the samples. A high representation around 1 kb was observed in all the samples. A gradual increase in smearing pattern was observed for the PCR products subjected to 18-27 number of PCR cycles. However, the smearing pattern observed for the PCR product after 24 cycles of PCR was above 5 kb, indicating the reactions were over-cycled (**Fig. 19**). Hence the number of PCR cycles for ds cDNA synthesis was optimized to 23.

To remove the redundant transcripts, 1 µg of ds cDNA was treated with 1 U, 0.5 U and 0.25 U of DSN. A control tube was maintained by adding DSN storage buffer. The treatment tubes were subjected to 7 cycles and then stored in ice, while the control tube was subjected to 7, 9, 11 and 13 PCR cycles and electrophorized. The optimum number of PCR cycles was selected as 11, since a smear appeared in high molecular weight region above 5 kb with PCR product from 13^{th} cycle (**Fig. 20**). The treatment tubes were then subjected to 4 more cycles to reach the optimum number of PCR cycles observed with the control template, followed by nine more PCR cycles as per the protocol. Among the three concentrations of DSN used, 0.25 U showed optimal normalization with reduced representation of transcripts around 1 kb, preserving the other transcripts. Gel electrophoresis indicated the disappearance of long PCR fragments when the concentration of DSN increased further (**Fig. 21**). qRT-PCR with *ACTIN* primers, a housekeeping gene was amplified from both normalized and non-normalized ds cDNA and normalization efficiency, ΔCq [ΔCq=Cq non-normalized cDNA- Cq normalized cDNA] was found to be 4.025 (**Fig. 22**).

4.3.1 Characterization of full length normalized cDNA library

When ds cDNA was ligated to pGEM-T Easy vector and cloned in *E. coli* JM109 (**Fig. 23**), 516 recombinants were obtained. Plasmid DNA was isolated from the recombinant clones and the insert region was amplified using PCR with M13 primers (Appendix 1) and electrophorized (**Fig. 24**). Four clones having insert size of 2.3 kb, one clone with insert size of 2.0 kb, 42 clones with insert size of 1.5-2 kb, 116 clones with insert size of range 1-1.5 kb and 353 clones with insert size below 1 kb were identified.

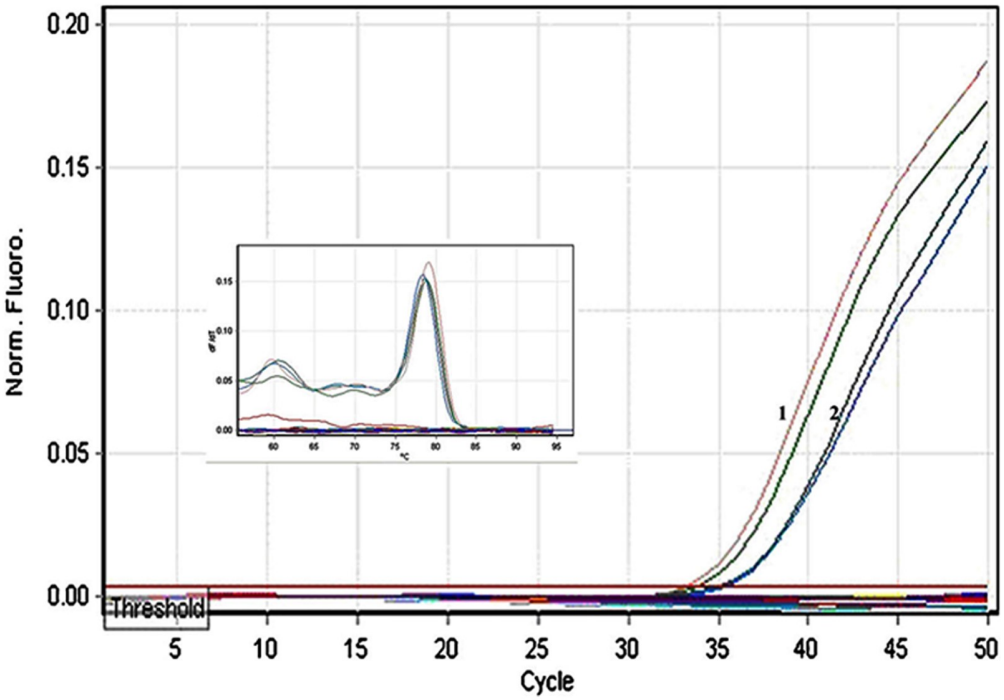


Figure 18. Amplification plot and melt curve of amplicon generated using mature miRNA sequences from total RNA of rhizome (1) and root (2) during qRT-PCR

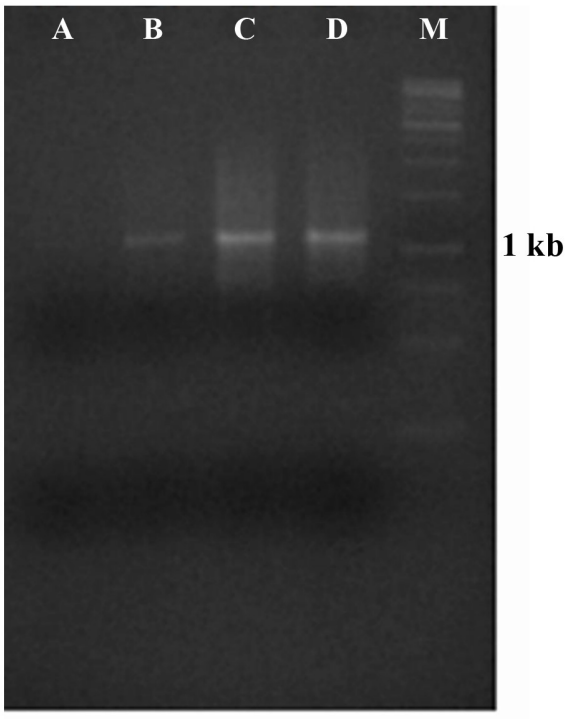


Figure 19. Agarose gel electrophoresis of PCR product when first strand cDNA was subjected to 18, 21, 24 and 27 number of PCR cycles represented in lanes A-D respectively. M-1 kb DNA marker

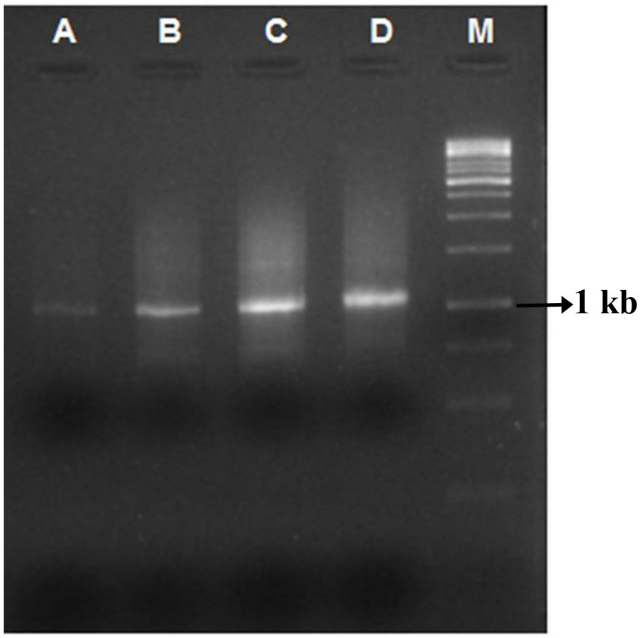


Figure 20. Agarose gel electrophoresis of PCR product when the double stranded cDNA was subjected to 7, 9, 11 & 13 number of PCR cycles represented from lane A-D respectively. M- 1 kb DNA marker

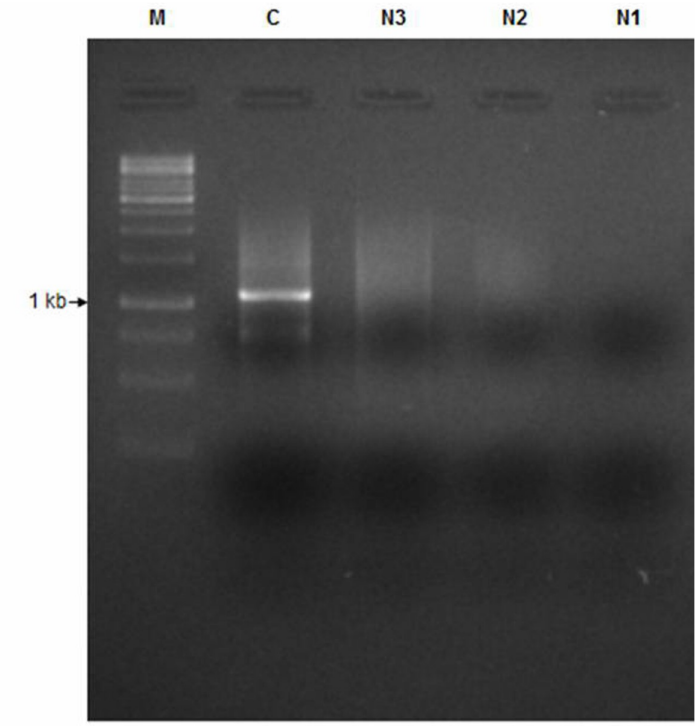


Figure 21. Agarose gel electrophoresis of double stranded cDNA after treating with 0.25 U (N3), 0.5 U (N2) and 1 U (N1) of DSN. C- control (without DSN). M- 1 kb DNA marker

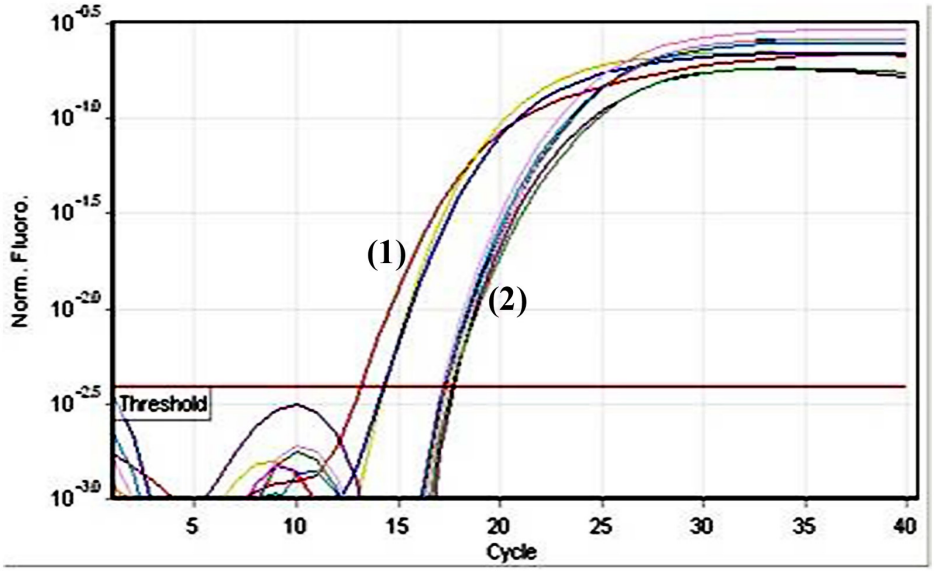


Figure 22. Amplification plot generated from control ds cDNA(1) and normalized ds cDNA (2) using actin primers

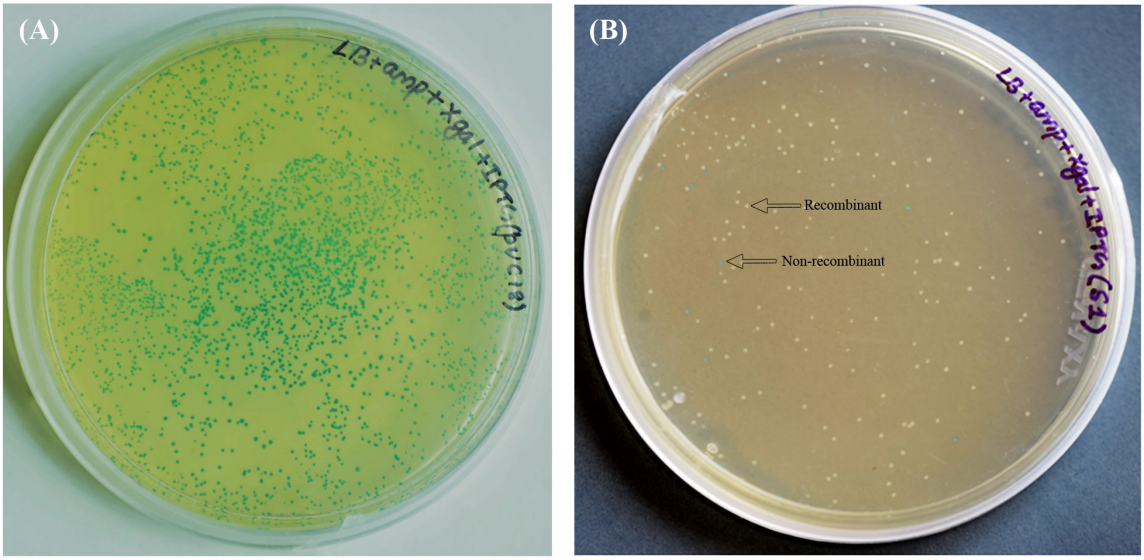


Figure 23. Representative plates showing blue-white screening of transformants (A) Control plate (B) Experimental plate

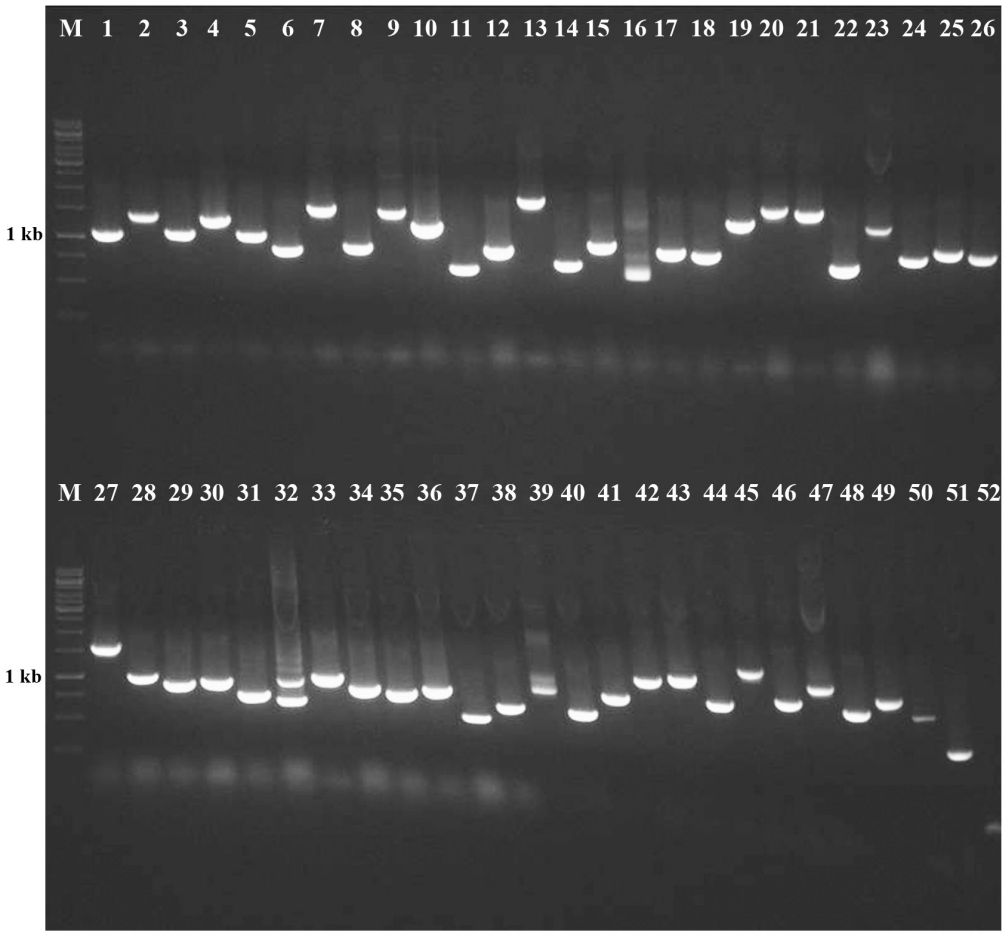


Figure 24. Representative agarose gel of PCR products of recombinant plasmids with M13 primers, M- 1 kb DNA marker.

Forty clones having insert size above 1.5 kb were sent for sequencing using M13 forward primer. However, only 16 clones gave sequencing result of >300bp which includes transcript showing similarity to MYB. The results are detailed in **Table 4**. Three clones showed no similarity when queried with non redundant protein sequences deposited in NCBI.

Table 4. Functional annotation of unigenes derived from normalized turmeric cDNA library

EST ID	Length (bp)	Gene description	Query coverage (%)	Identity (%)
NCL-EST1	288	MADS flowering locus C-like protein 2	15	100
NCL-EST2	572	Shaggy-related protein kinase	33	80
NCL-EST3	862	Arginine/ serine–rich splicing factor 2 variant 1	15	68
NCL-EST4	798	E3 ubiquitin–protein ligase RHF2A	57	78
NCL-EST5	625	Plasma membrane ATPase	86	98
NCL-EST6	823	26S protease regulatory subunit 6B homolog	75	98
NCL-EST7	592	Sister chromatid cohesion 1 protein 4 isoform X1	90	13
NCL-EST8	317	MADS flowering locus C-like protein 3	45	48
NCL-EST9	315	MYB13–*Triticum aestivum* (AtMYB59)	51	50
NCL-EST10	560	Eukaryotic translation initiation factor 3	64	91
NCL-EST11	865	Glutamate–rich WD repeat containing protein 1	58	76
NCL-EST12	515	No similarity	-	-
NCL-EST13	740	Uncharacterized protein	34	74
NCL-EST14	1125	No similarity	-	-
NCL-EST15	505	No similarity	-	-

4.4 Identification of candidate genes involved in curcumin biosynthesis

Transcriptomes of high-curcumin variety, Megha turmeric and *C. aromatica* (wild turmeric), which is practically devoid of curcumin, constructed from four month old rhizomes using Illumina HiSeq 2000 sequencing platform and deposited in NCBI with the BioProject ID PRJNA270561 and PRJNA277549 respectively, were used to mine the candidate genes of curcumin biosynthetic pathway. A total of 3.30- and 3.71-Gb paired-end short reads of 100 nt were generated from rhizomes of *C. longa and C. aromatica*, respectively. After removing the adapter and low-quality sequences from the raw data, 2.87-Gb (86 bp×2) and 3.16-Gb (85 bp×2) high-quality reads were obtained for *C. longa* and *C. aromatica* respectively. After assembly, 343144 transcripts and 381377 transcripts were identified for *C. longa* and *C. aromatica* respectively (Sheeja et al., 2015).

4.4.1 Mining of reported candidate genes of curcumin biosynthesis

When the transcriptomes were scanned for the reported key genes of curcuminoid biosynthesis using keyword search, all the ten reported genes namely phenylalanine ammonia lyase (*PAL*), cinnamate 4-hydroxylase (*C4H*), 4-coumarate-CoA ligase (*4CL*), hydroxy cinnamoyl-CoA transferase (*HCT*), coumarate 3-hydroxylase (*C3H*), caffeoyl CoA 3-O-methyltransferase (*COMT*), diketide CoA synthase (*DCS*), curcumin synthase 1 (*CURS1*), curcumin synthase 2 (*CURS2*) and curcumin synthase 3 (*CURS3*) were detected from both the transcriptomes and the number of unigenes detected for each gene was listed in **Table 5**. Polyketide synthases are reported as a key enzyme which forms the basic skeleton of flavonoids and various plant polyphenols (Schroder, 1997; Austin and Noel, 2003). Since curcumin is a polyphenol, the study was focussed to identify the key polyketide synthase involved in curcumin biosynthesis. Although the transcriptomes of high curcumin turmeric variety Megha turmeric was compared with *C. aromatica*, which is a zero curcumin genotype, no significant difference in the gene expression was detected for two of the important polyketide synthases ie., diketide CoA synthase and curcumin synthase.

Table 5. Statistics of putative genes involved in curcuminoid biosynthesis

Name of the gene	Number of unigenes	
	C. longa	***C. aromatica***
Phenylalanine ammonia lyase	3	9
Cinnamate 4-hydroxylase	8	9
4-Coumarate:CoA ligase	28	27
Coumarate 3-hydroxylase	2	1
Hydroxycinnamoyl-CoA transferase	5	7
Caffeoyl CoA O-methyltransferase	3	3
Diketide CoA synthase	3	3
Curcumin synthase 1	1	2
Curcumin synthase 2	2	3
Curcumin synthase 3	3	4

4.4.2 Unravelling *Curcuma* transcriptomes for novel polyketide synthases and transcription factors

The transcriptome was also scanned for the presence of polyketide synthases using local blast analysis in the Bioedit software (Hall, 1999) and keyword search. Twenty nine and thirty four *PKS* unigenes which include curcumin synthase, diketide CoA synthase and chalcone synthase were identified from the transcriptomes of *C. longa* and *C. aromatica* respectively (**Table 6a**, **6b**). DESeq analysis of *PKS* unigenes showed that twelve unigenes

were up-regulated and 10 unigenes were down-regulated in *C. longa* (**Table 7**). Functional annotation of twelve up-regulated *PKS* unigenes by gene ontology analysis annotated two unigenes as curcumin synthase (691222 and 652976) and ten unigenes as chalcone synthase (**Table 8**). Of the up-regulated transcripts, five transcripts (691222, 675683, 698710, 723003 and 674483) showed higher expression in *C. longa* with fold change >60, with the highest fold change of 69.25 observed in the 326 bp transcript (691222).

Table 6a. List of polyketide synthase genes identified from *C. aromatica*

Contig ID	Length (bp)	Gene Description	Identity (%)
691222	326	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	84
698710	350	Chalcone synthase [*Curcuma aromatica*]	100
723003	470	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	82
674483	284	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	76
599314	182	Chalcone synthase [Petunia x hybrida]	69
683784	306	Chalcone synthase [*Kaempferia elegans*]	94
608568	190	Chalcone synthase [*Kaempferia elegans*]	94
652976	244	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	77
662486	260	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	77
672845	281	Chalcone synthase [*Kaempferia elegans*]	82
625275	207	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	77
618431	199	NA	98
627366	209	Curcumin synthase 1	100
572607	161	NA	100
578094	165	Curcumin synthase 3	100
704218	371	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	68
607258	188	Curcumin synthase 3	98
606629	188	Curcumin synthase 2	100
636441	220	Curcumin synthase 2	100
628472	210	Curcumin synthase 3	100
684812	309	Curcumin synthase 3	98
684236	307	Curcumin synthase 2	100
730066	525	NA	100
616477	197	NA	100
557786	152	NA	100
581526	168	Chalcone synthase [*Kaempferia elegans*]	85
563944	156	Chalcone synthase [*Boesenbergia rotunda*]	97
631627	214	NA	100
623945	205	Chalcone synthase [*Boesenbergia rotunda*]	93
586908	172	Chalcone synthase [*Boesenbergia rotunda*]	98
721357	458	Chalcone synthase [*Boesenbergia rotunda*]	98
729218	518	NA	99
720925	455	NA	100

Table 6b. List of polyketide synthase genes identified from *C. longa*

Contig ID	Length (bp)	Gene Description	Identity (%)
675683	962	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	79
562074	216	Chalcone synthase [*Kaempferia elegans*]	96
535759	187	Polyketide synthase type III isoform 1 [*Wachendorfia thyrsiflora*]	82
648985	482	Curcumin synthase 3 [*C. longa*]	81
545638	197	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	86
538092	189	Chalcone synthase [*Kaempferia elegans*]	82
663071	635	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	78
643929	447	Diketide CoA synthase [*C. longa*]	98
639158	418	Curcumin synthase 1 [*C. longa*]	100
614566	319	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	91
584369	250	Curcumin synthase 3 [*C. longa*]	100
497761	157	Curcumin synthase 2 [*C. longa*]	100
502863	160	Curcumin synthase 2 [*C. longa*]	100
623284	348	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	73
680765	1274	Polyketide synthase [*Theobroma cacao*]	72
609988	306	Curcumin synthase 3 [*C. longa*]	99
571905	230	NA	100
551543	203	NA	100
654444	529	NA	100
517000	171	NA	100
507786	164	Chalcone synthase [*Kaempferia elegans*]	85
589408	259	Chalcone synthase [*Boesenbergia rotunda*]	89
489206	151	Chalcone synthase [*Boesenbergia rotunda*]	91
636658	404	Chalcone synthase [*Boesenbergia rotunda*]	96
644993	453	NA	99
630392	376	NA	99
581611	245	NA	100
520843	174	NA	98
638348	413	NA	100

Table 7. List of differentially expressed polyketide synthase genes

Contig ID	Gene description	Identity (%)	Fold change in *C. longa* compared with *C. aromatica*
691222	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	84	69.25
675683	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	79	68.17
698710	Chalcone synthase [*Curcuma aromatica*]	100	66.96
723003	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	82	65.85
674483	Type III polyketide synthase 2 [*Musa acuminata* AAA Group]	76	63.96
562074	Chalcone synthase [*Kaempferia elegans*]	96	36.82
599314	Chalcone synthase [Petunia x hybrida]	69	35.05
683784	Chalcone synthase [*Kaempferia elegans*]	94	32.84
535759	Polyketide synthase type III isoform 1 [Wachendorfia *thyrsiflora*]	82	29.80
608568	Chalcone synthase [*Kaempferia elegans*]	94	20.98
648985	Curcumin synthase 3 [*C. longa*]	81	20.82
652976	Type III polyketide synthase 4 [*Musa acuminata* AAA Group]	77	17.10
636658	Chalcone synthase [*Boesenbergia rotunda*]	96	0.08
586908	Chalcone synthase [*Boesenbergia rotunda*]	98	0.08
721357	Chalcone synthase [*Boesenbergia rotunda*]	98	0.07
644993	ClPKS7 [*C. longa*]	99	0.03
630392	ClPKS9 [*C. longa*]	99	0.01
581611	ClPKS9 [*C. longa*]	100	0.01
729218	ClPKS9 [*C. longa*]	99	0.01
520843	ClPKS9 [*C. longa*]	98	0.01
638348	ClPKS9 [*C. longa*]	100	0.01

Table 8. Gene ontology analysis of up-regulated polyketide synthase transcripts in *C. longa*

Contig ID	Length (bp)	Enzyme code	Sequence description	GO names	InterPro motif details
691222	326	EC:2.3.1.74	curcumin synthase	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN
698710	350	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, biosynthetic process	IPR016039 Thiolase-like DOMAIN
723003	470	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN
674483	284	EC:2.3.1.74	chalcone synthase	transferase activity, transferring acyl groups other than amino-acyl groups, protein homodimerization activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN

599314	182	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN
683784	306	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN
608568	190	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, biosynthetic process	IPR001099 Chalcone/stilbene synthase, N-terminal DOMAIN
652976	244	EC:2.3.1.74	curcumin synthase partial	transferase activity, transferring acyl groups other than amino-acyl groups, biosynthetic process	IPR016039 Thiolase-like DOMAIN
675683	962	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process	IPR012328 Chalcone/stilbene synthase, C-terminal DOMAIN
562074	216	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN
535759	187	EC:2.3.1.74	putatative chalcone synthase	naringenin-chalcone synthase activity, biosynthetic process	-
648985	482	EC:2.3.1.74	chalcone synthase	naringenin-chalcone synthase activity, flavonoid biosynthetic process	IPR016039 Thiolase-like DOMAIN

Alignment of the annotated transcripts of *C. longa* to AGRIS database resulted in the identification of unigenes belonging to 39 transcription factor families in *C. longa* and 36 transcription factor families in *C. aromatica* (**Fig. 25**). The members of AP2-EREBP, MYB, C2H2, C2C2-Dof, WRKY, bHLH, HOMEOBOX, TCP, NAC and bHLH families were dominant in both the species. Four hundred and six unigenes in *C. longa* showed similarity to R2R3-MYB. Of which, 18 transcripts showed higher expression (>2 fold) in *C. longa*, while 41 transcripts showed higher expression in *C. aromatica*.

4.5. Identification of stable reference genes in turmeric

Accurate quantification of expression of genes relies on the selection of stable reference genes used to normalize the candidate gene expression under different experimental conditions. The expression of six genes namely, *EF1α, UBIQUITIN, 18S rRNA, GAPDH, ACTIN* and *TUBULIN* were analyzed by qRT-PCR using cDNA synthesized from leaves, pseudostem, rhizomes and roots across three developmental stages (60, 120 and 180 DAP) to determine their stability. The primers for the amplification of *EF1α, UBIQUITIN, 18S rRNA, GAPDH, ACTIN* and *TUBULIN* are listed

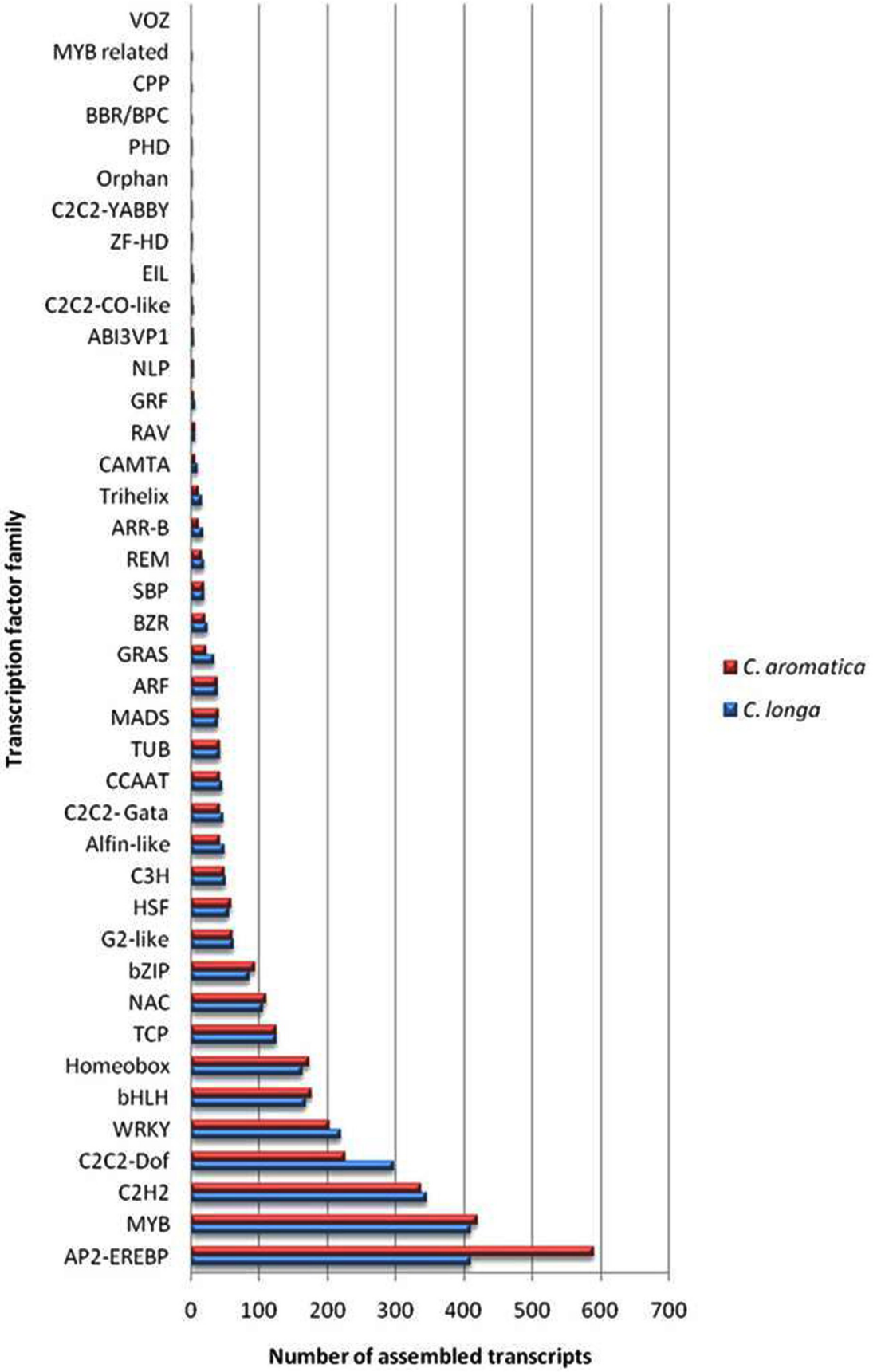


Figure 25. Distribution of different transcription factors in the assembled transcripts of *C. longa* and *C. aromatica*

in Appendix 1. Gene specific amplification was confirmed by visualizing a specific amplicon in 2% agarose gel electrophoresis (**Fig. 26**), presence of a single peak in dissociation curve analysis (**Fig. 27**) and by sequencing. The PCR efficiency for the reference genes ranged from 92% to 116% (**Table 9**) with 99% for *18S rRNA* and 101% for *EF1α* and *ACTIN*. The regression coefficient (r^2) value was 0.970 to 0.996, with highest for *UBIQUITIN* (0.996) and least for *18S rRNA* (0.970). The stability of reference genes was analyzed using Reffinder. According to ranking by Delta CT method (**Fig. 28**) and Normfinder (**Fig. 29**), *EF1α* was the best gene. Genorm identified *EF1α* and *UBIQUITIN* as the stable pair genes with M value of 1.025. *EF1α, UBIQUITIN, 18S rRNA* and *GAPDH* had Genorm M value below 1.5 (**Fig. 30**). Bestkeeper identified *18S rRNA* as the stable gene (**Fig. 31**). The comprehensive ranking in the Reffinder was in the order: *EF1α <UBIQUITIN <18S rRNA <GAPDH <ACTIN <TUBULIN* (**Table 10**). Hence, following the comprehensive ranking, the most stable genes, *EF1α* and *UBIQUITIN* were selected as reference genes for normalizing the expression of target gene.

Table 9. Characteristics of calibration curves in the amplification of reference genes

	Regression coefficient (r^2)	**PCR efficiency (%)**
EF1α	0.988	101
UBIQUITIN	0.996	116
ACTIN	0.992	101
18S rRNA	0.970	99
GAPDH	0.998	109
TUBULIN	0.989	92

Table 10. Stability of reference gene expression in developing turmeric tissues

Ranking Order (Better--Good--Average)						
Method	1	2	3	4	5	6
Delta CT	*EFIα*	*UBIQUITIN*	*18S rRNA*	*GAPDH*	*ACTIN*	*TUBULIN*
BestKeeper	*18S rRNA*	*EFIα*	*ACTIN*	*UBIQUITIN*	*GAPDH*	*TUBULIN*
Normfinder	*EFIα*	*UBIQUITIN*	*18S rRNA*	*GAPDH*	*TUBULIN*	*ACTIN*
Genorm	*EFIα / UBIQUITIN*		*18S rRNA*	*GAPDH*	*ACTIN*	*TUBULIN*
Recommended comprehensive ranking	*EFIα*	*UBIQUITIN*	*18S rRNA*	*GAPDH*	*ACTIN*	*TUBULIN*

4.6 Expression analysis of candidate genes involved in curcumin biosynthesis

The accumulation of secondary metabolite was reported to be influenced by a number of factors like genotype, tissue type, developmental stage, environment which

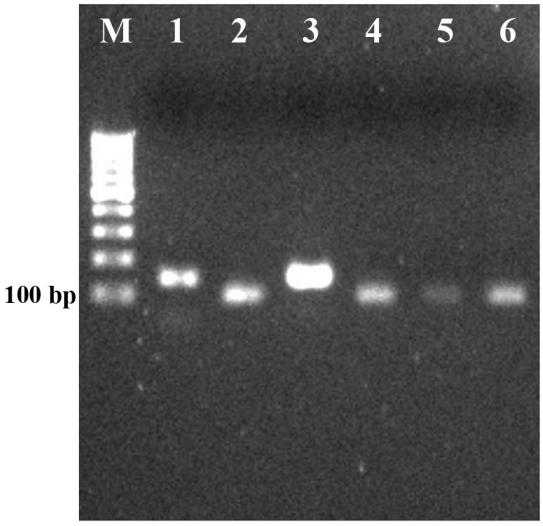


Figure 26. Agarose gel electrophoresis of real time PCR products of candidate reference genes. M- 100 bp DNA marker, Lane 1-6 shows specific amplification products of *EF1α, UBIQUITIN, ACTIN, 18S rRNA, GAPDH* and *TUBULIN*

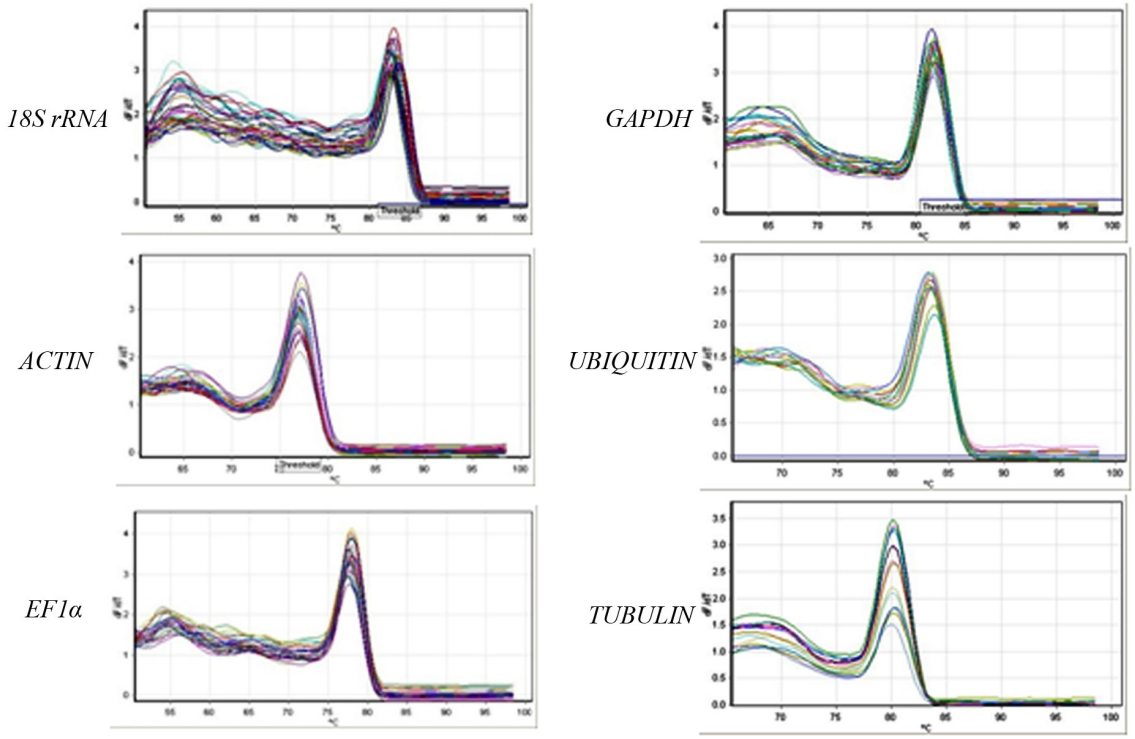


Figure 27. Melt curve analysis of reference genes

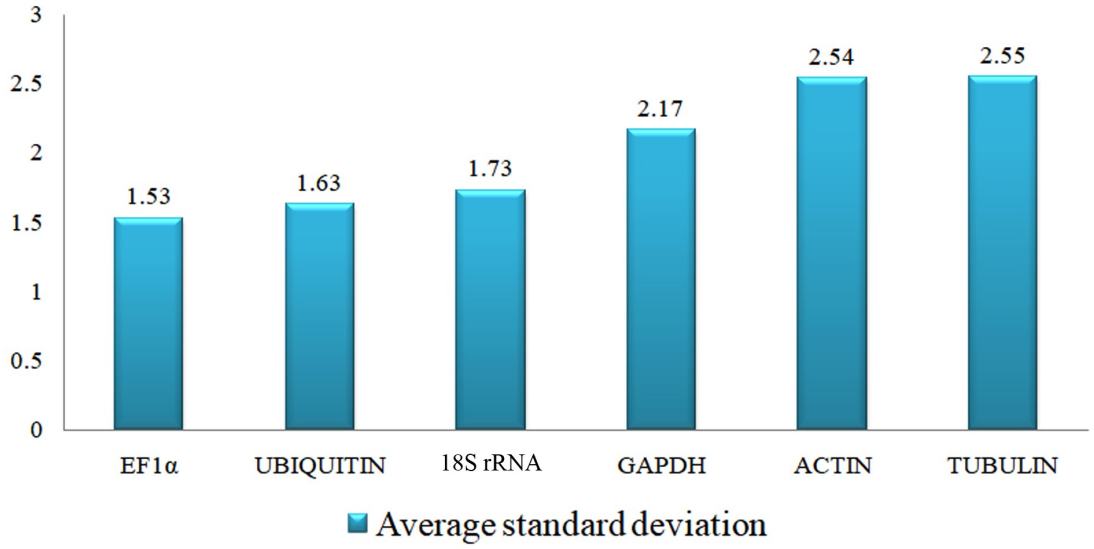


Figure 28. Average standard deviation of candidate reference genes as determined by Delta Ct method

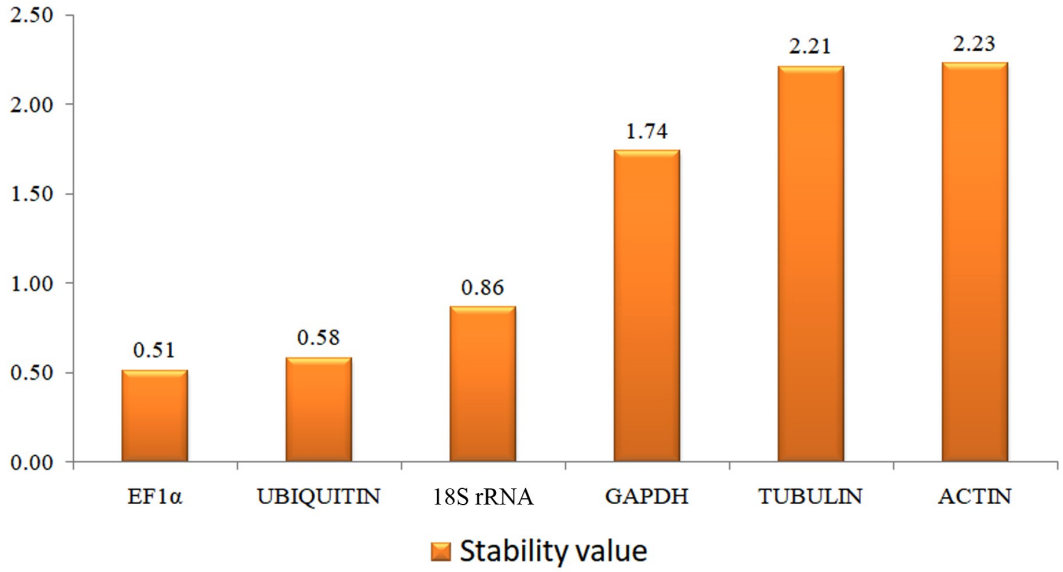


Figure 29. Average expression stability values of candidate reference genes as determined by NormFinder

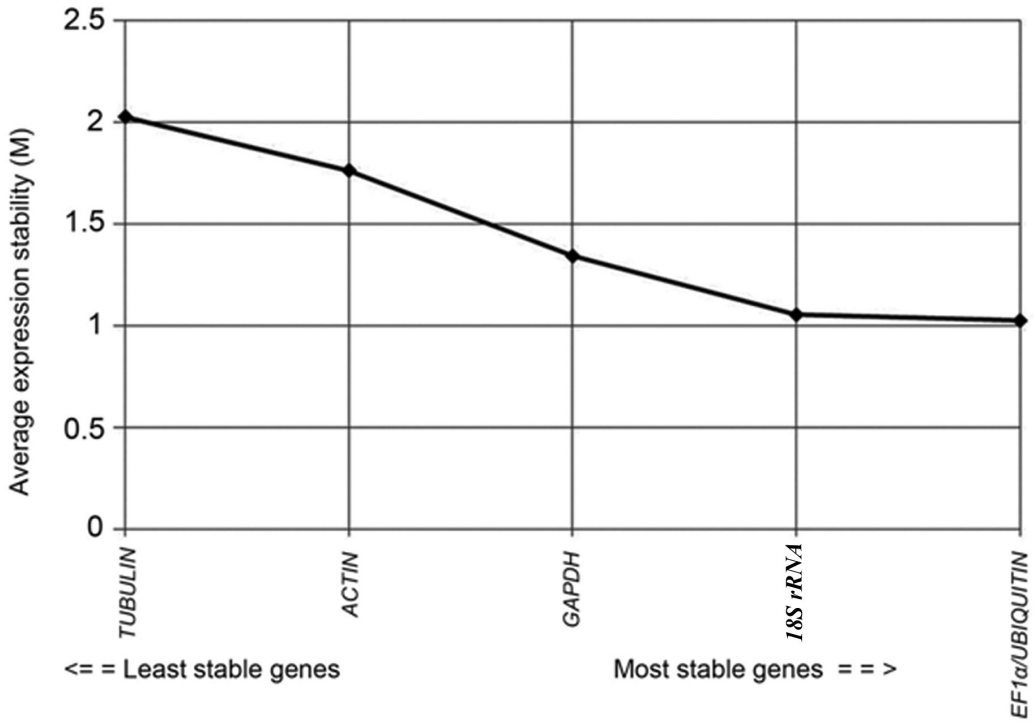


Figure 30. Average expression stability values of candidate reference genes as determined by GeNorm

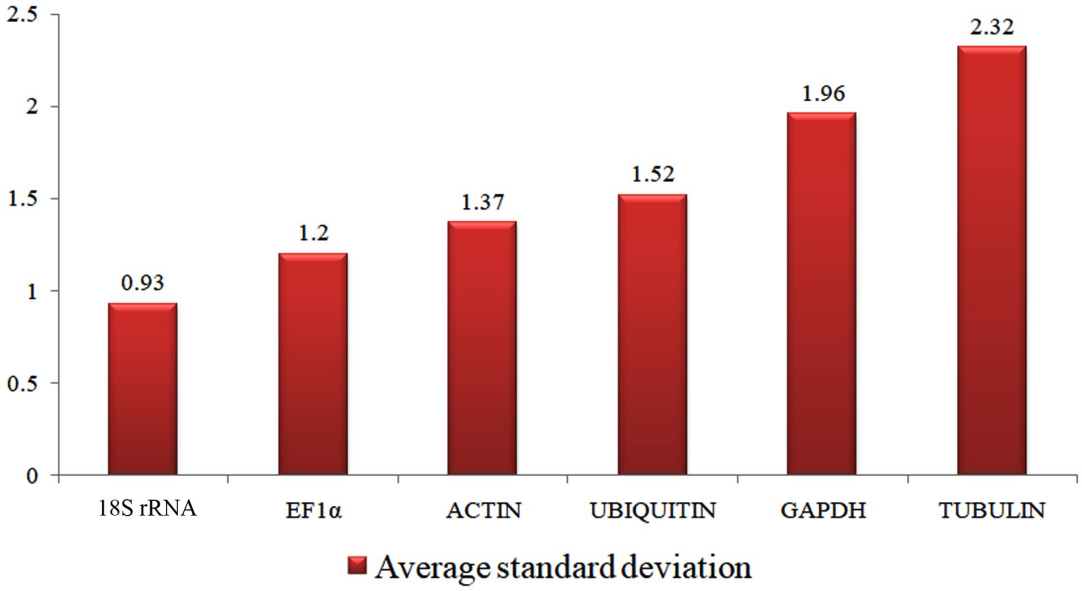


Figure 31. Average standard deviation of candidate reference genes as determined by Bestkeeper

includes various biotic and abiotic stresses (Cultrone et al., 2010; Smita et al., 2013; Zandalinas et al., 2016; Jeena et al., 2017; Wentzell and Kliebenstein, 2008). Hence, to explore the impact of environment, genotype, tissue type and developmental stage on curcumin accumulation and to correlate metabolite (curcumin) content with gene expression, two genotypes of turmeric with contrasting curcumin contents were studied under three different experimental conditions.

Tissue and developmental stage specific expression of ten reported candidate genes and eleven transcripts showing differential expression were analyzed in leaves and rhizomes of high curcumin variety, IISR Prathibha at 60, 120 and 180 DAP. The sequences of gene specific primers are detailed in Appendix 1. In leaves, the expression of *PAL, 4CL, C3H, HCT, COMT, CURS3, ClPKS d-g, MYB1, MYB2* and *bHLH* was highest at 60 DAP. At 120 DAP, *DCS, CURS1, CURS2* and *WRKY* showed higher expression and the transcript level of *C4H, ClPKS a-c* was higher at 180 DAP (**Fig. 32**). In rhizomes, the expression of *PAL, C3H, HCT, COMT*, *MYB1* and *bHLH* was highest at 60 DAP and that of *DCS, CURS1, CURS2* and *WRKY* were higher at 120 DAP, which is similar to that observed in leaves. The transcript level of *C4H, 4CL, CURS3* and *ClPKS a-g* were also observed to be higher at 120 DAP in rhizomes, when the curcumin was at its maximum (6.52%). At 180 DAP, only the expression of *MYB2* was found to be up-regulated (**Fig. 33**). When both these tissues were compared (**Fig. 34**), the expression of *C4H, C3H, HCT, COMT, CURS1, CURS3, ClPKSa-g,* MYB and WRKY was higher at 120 DAP in rhizomes.

To explore the influence of genotypic and environmental factors on gene expression, twenty one transcripts encompassing the above mentioned category of genes were assessed by qRT-PCR in IISR Prathibha and Collection no. 449, under three different environments namely under field conditions at Kozhikode (Koz-F) (11.2994°N, 75.8407°N), under green house conditions at Kozhikode (Koz-GH) and under field conditions at Coimbatore (Cbe-F) (9.1913°N, 77.8803°E). The average relative humidity, annual rainfall and temperature of Kozhikode are 85%, 3205 mm and 27.3 ºC whereas that of Coimbatore is 61%, 618 mm and 26.3 ºC respectively. Under field conditions, standard package of practices was followed. In green house condition, nutrient stress was imposed by depriving fertilizers/ input application.

The expression of *PAL, C4H, 4CL, C3H, COMT, DCS, CURS1, CURS3, ClPKS a-d,* WRKY and bHLH were up-regulated in IISR Prathibha compared to collection no. 449. However, *HCT, CURS2 and ClPKS e-g* were down-regulated in IISR Prathibha in Koz-GH. In Koz-F, the expression of two MYB transcripts was also found to be lower in IISR Prathibha (**Fig. 35**). When the expression of IISR Prathibha alone was compared in three environmental conditions, the expression of *PAL, C4H, 4CL, C3H, COMT, DCS, CURS1-3* and *ClPKS a-g* was prominent at Koz-F, where the maximum curcumin (6.52%) accumulation was observed. However, the expression of *HCT, MYB* and *bHLH* was higher in Koz-GH. The expression of *MYB2* was found to be higher in both Koz-GH and Cbe-F (**Fig. 36**).

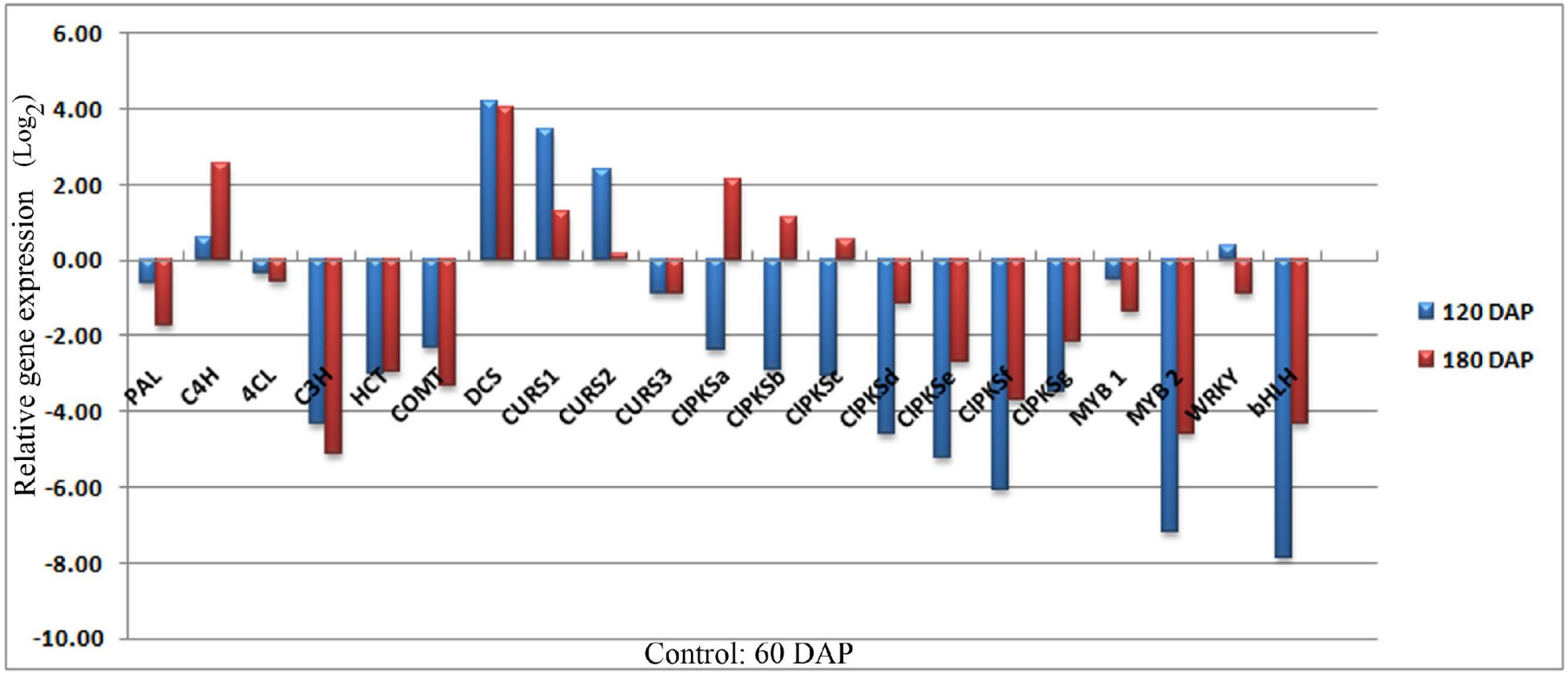


Figure 32. Relative gene expression analysis of selected genes in leaves of turmeric harvested at 60, 120 and 180 DAP

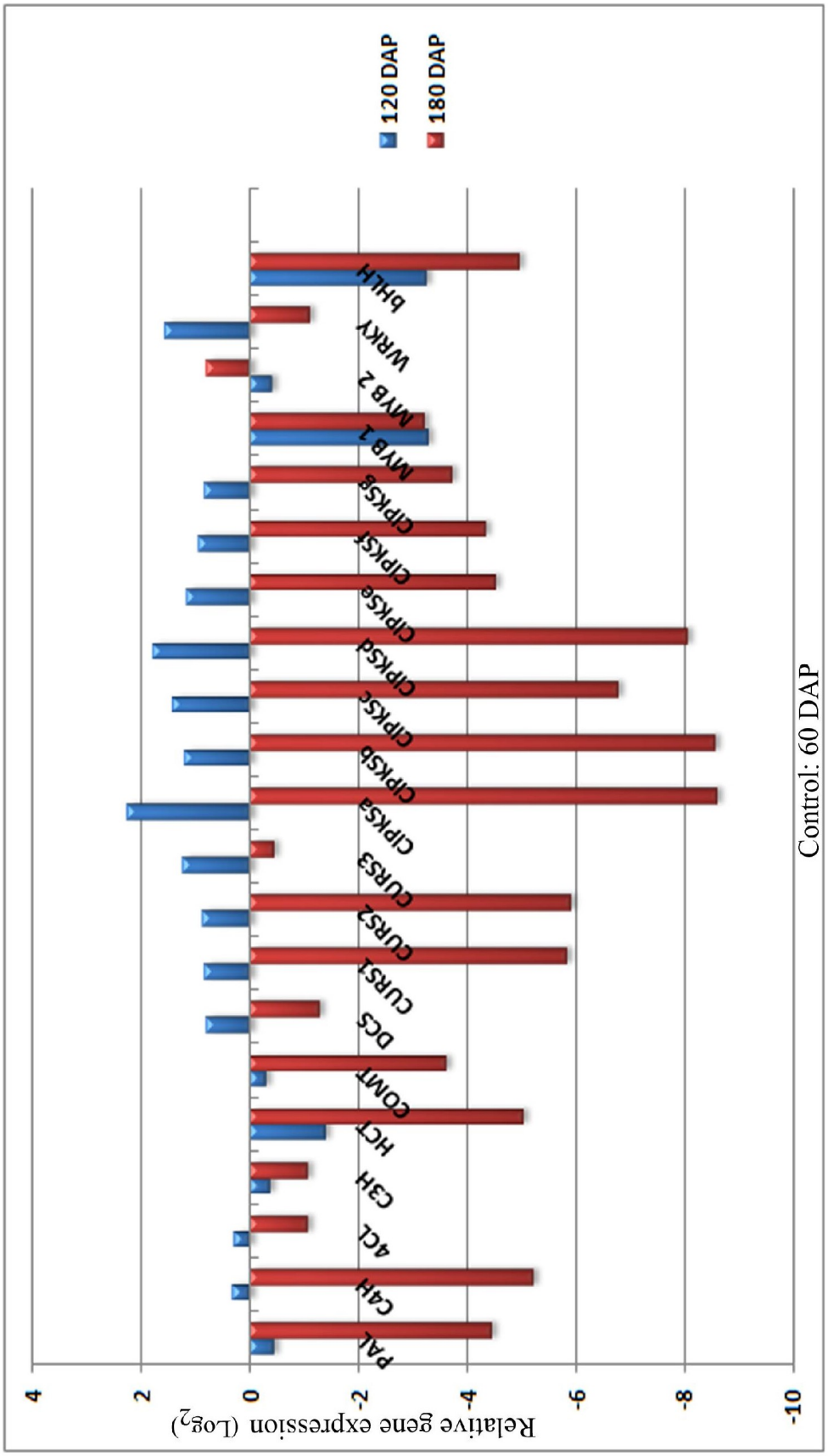


Figure 33. Relative gene expression analysis of selected genes in rhizomes of turmeric harvested at 60, 120 and 180 DAP

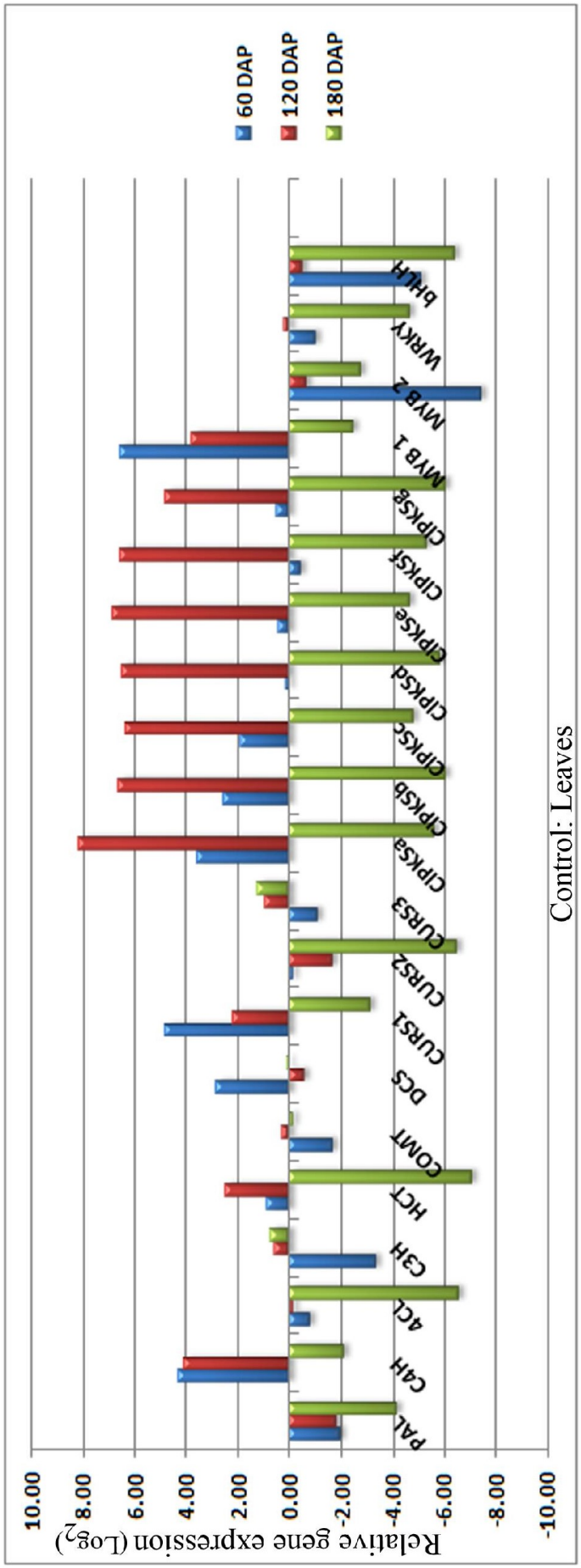


Figure 34. Comparison of relative gene expression analysis of selected genes in rhizomes vs leaves of turmeric harvested at 60, 120 and 180 DAP

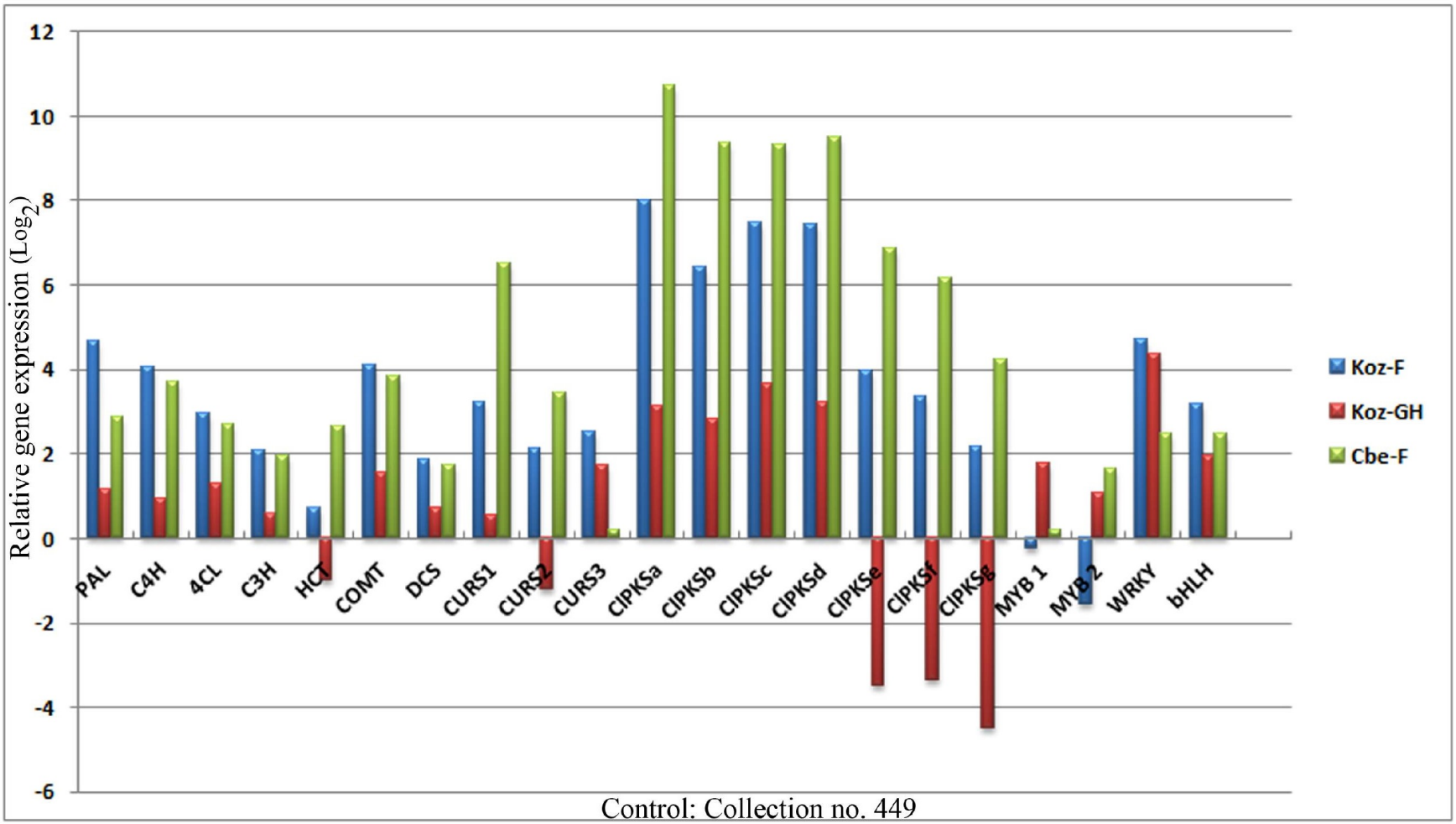


Figure 35. Comparison of relative gene expression analysis of selected genes in rhizomes of IISR Prathibha vs Collection no. 449 harvested from three environmental conditions

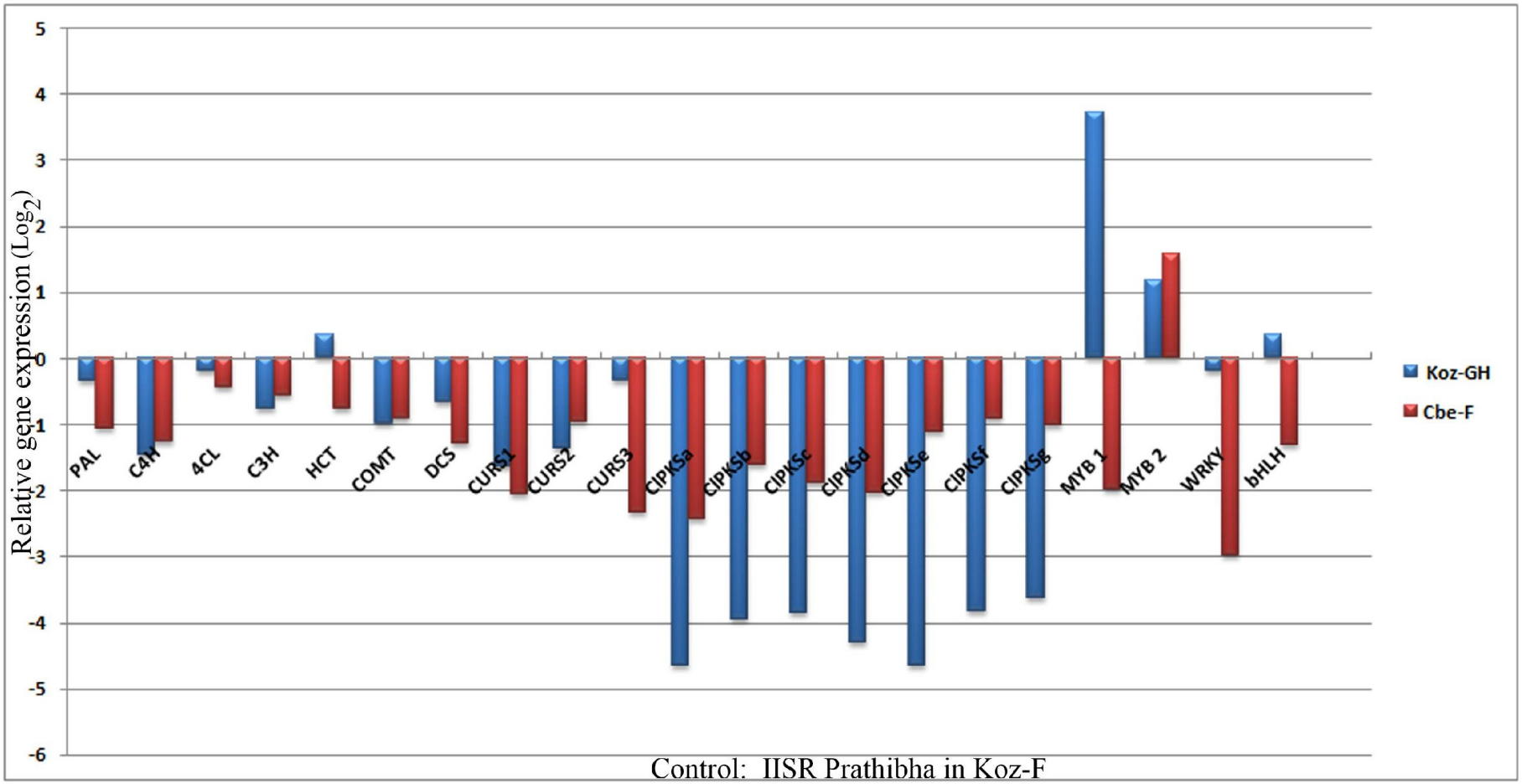


Figure 36. Relative gene expression analysis of selected genes in rhizomes of IISR Prathibha harvested from three environmental conditions

4.7 Estimation of curcumin content and its correlation with normalized gene expression

The curcumin content of IISR Prathibha as determined by ASTA method (1968) with ethanol as solvent was found to be maximum in rhizomes when harvested at 120 DAP (6.52%) compared to rhizomes harvested at 60 dap (0.79%) and 180 dap (5.73%). Similarly, the curcumin content of IISR Prathibha was higher in Koz-F (6.52%) than Cbe-F (4.53%) and was lowest under nutrient limiting condition in Koz-GH (1.54%). Similar trend was observed in Collection no. 449, which showed highest curcumin content in Koz-F (2.11%) compared with Cbe-F (0.56%) with negligible curcumin content under nutrient limiting conditions in Koz-GH (0.23%). The correlation analysis between curcumin accumulation and the normalized gene expression was performed using Pearson's correlation coefficient. The rhizomes of IISR Prathibha and Collection no. 449 of 120 DAP planted under the three environmental conditions mentioned above were used for the study. Correlations are considered positive when Pearson's r>0.6 (Wang et al., 2015). There was a direct correlation between the transcript levels of *ClPKSa, ClPKSb, ClPKSc, ClPKSd, ClPKSe, C4H, 4CL, C3H, COMT*, *DCS* and *WRKY* and the accumulation of curcumin. Among the twenty one transcripts analyzed, transcripts *ClPKSd* showed highest positive correlation (r=0.8742) and *MYB2* showed the least (r=0.0979) (**Table 11**).

Table 11. Pearson's correlation of the curcumin content with the gene expression when different tissues were analyzed at three developmental stages from two turmeric genotypes planted in three environmental conditions

Transcript	**Pearson's *r* value**	**Transcript**	**Pearson's *r* value**
ClPKSd	0.8742	*CURS1*	0.5994
ClPKSb	0.8662	*ClPKSf*	0.5807
ClPKSc	0.8646	*PAL*	0.5055
ClPKSa	0.8514	*CURS2*	0.4955
C3H	0.7622	*bHLH*	0.4925
C4H	0.7304	*CURS3*	0.4882
COMT	0.7136	*ClPKSg*	0.4053
DCS	0.6668	*MYB2*	-0.2868
ClPKSe	0.6393	*HCT*	0.1117
WRKY	0.6298	*MYB1*	0.0979
4CL	0.6181		

4.8 Full length amplification of key genes involved in curcumin biosynthesis

4.8.1 Amplification of coding region of reported *PKSs* of curcumin biosynthesis

To characterize a transcript, it is essential to amplify its full length coding region. The full length coding region of *DCS, CURS1*and *CURS3* were PCR amplified from DNA of IISR Prathibha (**Fig. 37**) using the primers *DCS-K, CURS1-K* and *CURS3-K* (Appendix 1) described in Katsuyama et al. (2009a and 2009b). The amplicon size of *DCS, CURS1* and *CURS3* were 1.4 kb, 1.2 kb and 1.3 kb respectively. These sequences were analyzed for the presence of intron using Genscan. The results showed that *CURS1* and *CURS3* had an intron of 66 bp and 88 bp respectively. *DCS* had two introns of 113 bp and 85 bp. Schematic diagrams of the positions of exons and introns in these genes are represented in **Fig. 38**.

4.8.2 Full length cDNA amplification of *CURS3*

The inverse PCR protocol combined with SMARTer (Switching Mechanism At 5' end of RNA Transcript) PCR technology was opted for the amplification of full length cDNA. To confirm the suitability of the protocol, *CURS3* was selected as a positive control as its full length coding region was already sequenced from gDNA as mentioned in section 4.8.1 and it is reported to have wider substrate utilization efficiency in comparison with *DCS, CURS1* and *CURS2* (Katsuyama et al., 2009b). Inverse PCR with circularized normalized double stranded cDNA and outward primers, *CURS3-IP* (Appendix 1) designed from the coding region of *CURS3* amplified ~1.1 kb fragment (**Fig. 39**). The PCR product was purified, ligated in pGEM-T easy vector, cloned in *E. coli* JM109 and three recombinant plasmids were sequenced using M13 primers. The sequences were aligned with coding sequence of *CURS3* and the 3' UTR and 5' UTR regions of size 150 bp and 127 bp respectively were identified. A putative uORF of 66 bp and an internal ribosome entry site from 5' UTR and a polyadenylation sequence from 3' UTR were predicted using UTRScan. A set of primers *CURS3-FL* (Appendix 1) was designed from UTR regions to confirm the sequence and was tested using PCR in three high curcumin turmeric genotypes, three low curcumin turmeric genotypes and two other *Curcuma* species (*C. aromatica* and *C. caesia*). An amplicon of ~1.2 kb fragment was observed in all the collections of *C. longa* irrespective of their curcumin content with highly conserved amino acid sequence in the coding region of *CURS3*. The amplification was not observed in other

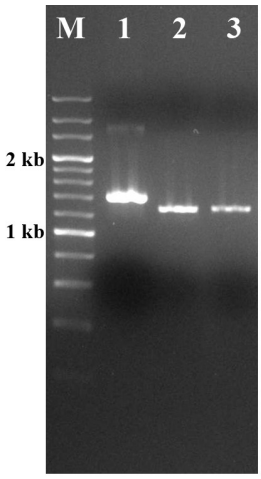


Figure 37. Amplification of full length ORF of reported PKSs from genomic DNA. Lane 1-3: PCR products of *DCS, CURS1* and *CURS3* respectively. M: 200 bp DNA marker

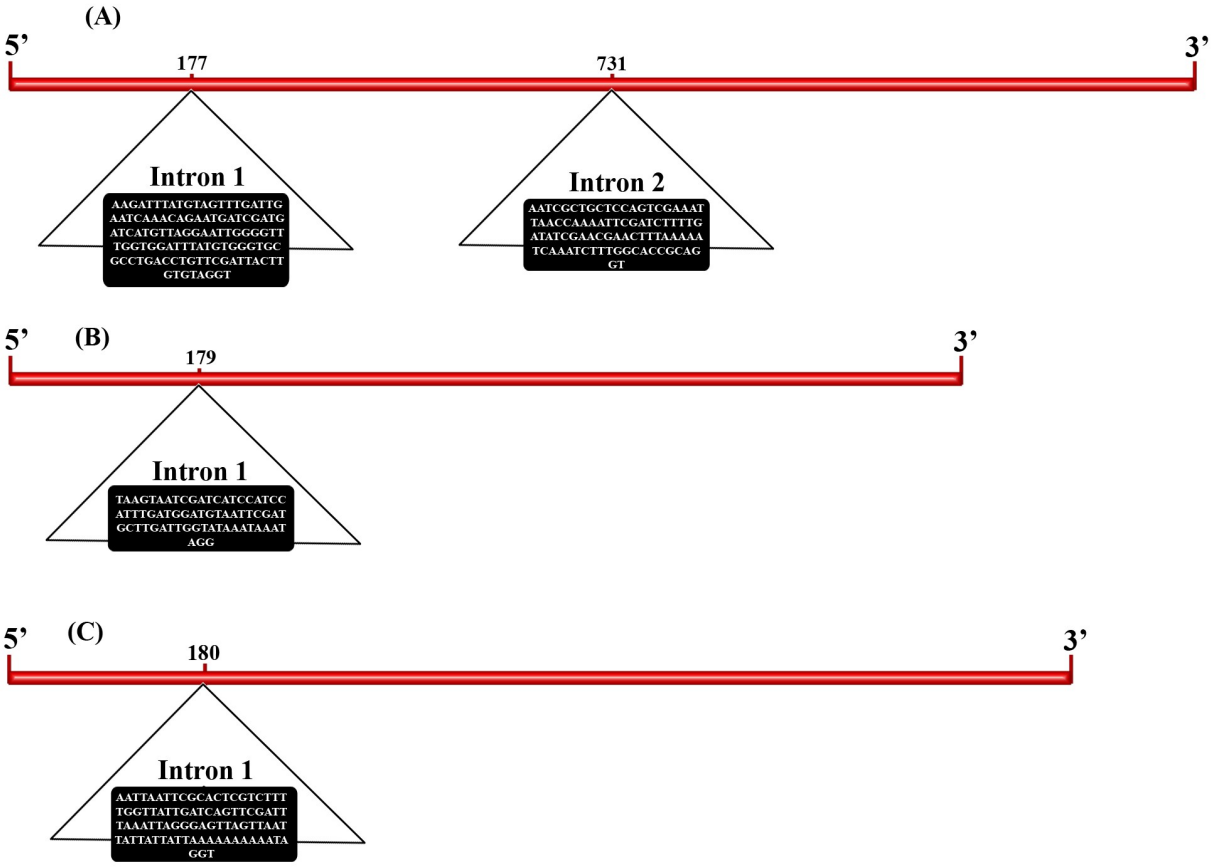


Figure 38. Schematic representation of introns in (A) *DCS* (B) *CURS1* and (C) *CURS3*

Curcuma species (*C. aromatica* and *C. caesia*) with these primers designed from UTR regions of *CURS3*. However, the presence of full length coding region of *CURS3* in *C. aromatica* and *C. caesia* was confirmed by PCR amplification using the primer (CURS3-K) as described in Katsuyama et al. (2009b). To confirm the UTR sequence variations between species, ds cDNA was synthesized from *C. aromatica*, self circularized and inverse PCR was carried out. The PCR product was sequenced and when aligned with *CURS3* of *C. longa*, sequence variations in the UTR regions were observed between the species (**Fig. 40**).

4.8.3 Full length cDNA amplification of novel *ClPKSa*

As described in section 4.7, five novel PKS transcripts (*ClPKSa-e*) showed positive correlation with curcumin content and the transcript with maximum relative gene expression (*ClPKSa*) was chosen for a detailed study. Inverse PCR using double stranded circularized cDNA with outward primers, *ClPKS11-IPa* (Appendix 1) and nested PCR with *ClPKS11-IPb* (Appendix 1) amplified an intense ~1.6 kb fragment. Re-amplification of purified inverse PCR product with primer set 1 [adaptor primer (PCR primer M1) - *ClPKS11* IPa-FP] / primer set 2 [adaptor primer- *ClPKS11* IPa-RP] was done and fragments of ~500 bp and ~1000 bp were amplified respectively (**Fig. 41**). The PCR product was purified, ligated in pGEM-T Easy vector, cloned in *E. coli* JM109 and three recombinant plasmids were sequenced using M13 primers. The sequence was aligned with the short sequence of *ClPKSa*, hereafter referred as *ClPKS11*. Apart from *ClPKSa, ClPKSb-d* also showed >96% identity with *ClPKS11*. Finally, an end to end PCR including both start and stop codon of *ClPKS11* was amplified from both cDNA (GenBank accession number-KX017475) and genomic DNA of *C. longa* (**Fig. 42**) using the primers *ClPKS11-FL* (Appendix 1). The PCR products were cloned and sequenced.

4.9 Structural characterization of *ClPKS11*

4.9.1 Sequence analysis of *ClPKS11*

ORF finder indicated that *ClPKS11* has a coding region of 1176 bp that corresponded to a deduced protein sequence of 391 amino acid residues (**Fig. 43**). The ORF region was bordered by 81 bp 5'-untranslated region (UTR) and 262 bp 3'-UTR. A putative uORF of 69 bp and a polyadenylation sequence were predicted from 3' UTR using

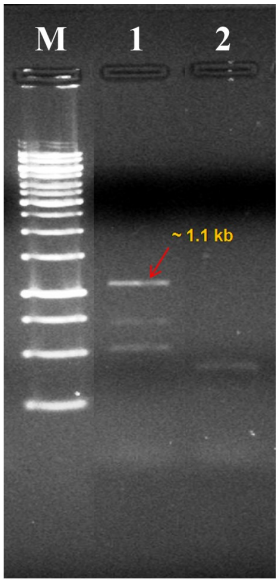


Figure 39. Agarose gel showing the inverse PCR product of ~1.1 kb amplified with *CURS3* IP primers (Lane 1) from circularized cDNA template but not with non-circularized cDNA (Lane 2). M- 1 kb DNA marker

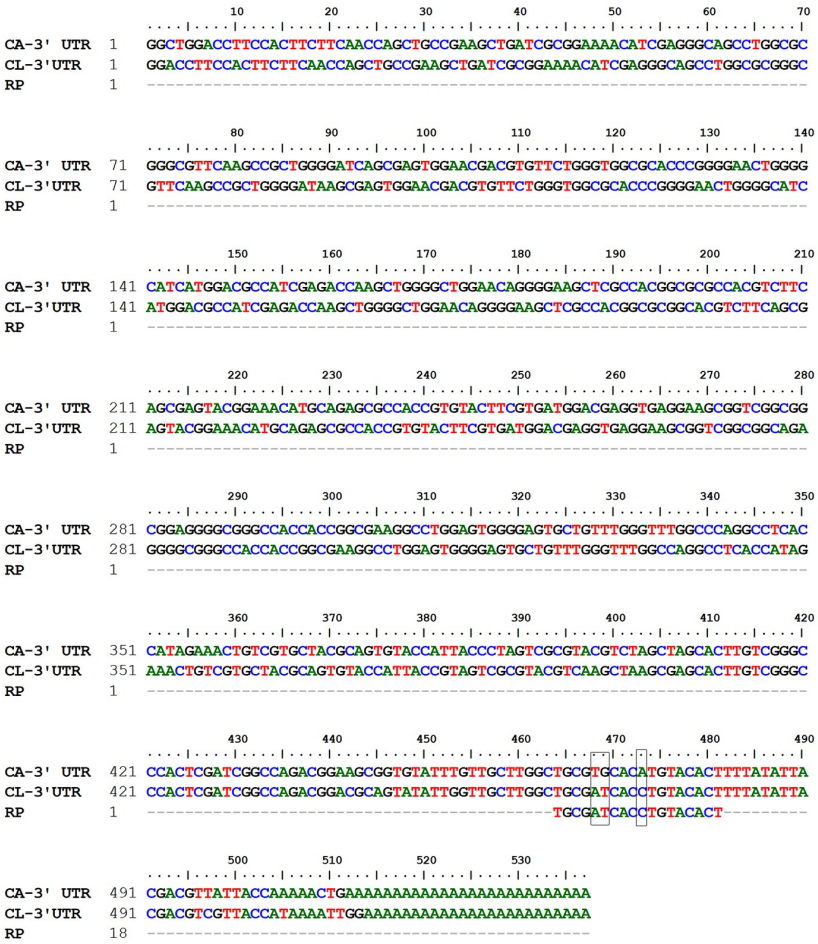


Figure 40. Comparison of 3' UTR regions of *CURS3* in *C. longa* and *C. aromatica*

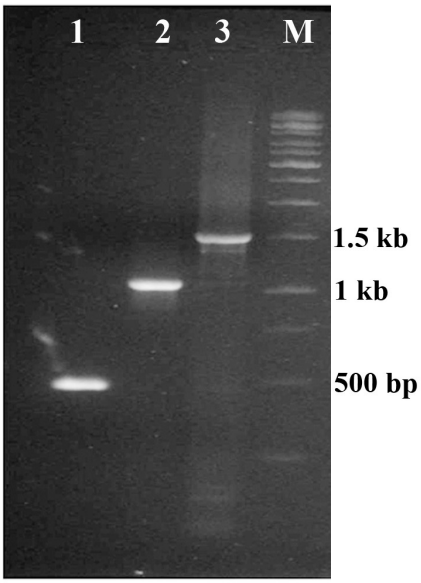


Figure 41. Agarose gel showing the inverse PCR product of ~1.6 kb amplified with *ClPKS11* IP primers (Lane 3) from circularized cDNA template. Lane 1 and 2 shows the specific amplification with PCR primer M1- *ClPKS11* IPa-FP and PCR primer M1- *ClPKS11* IPa-RP respectively. M- 1 kb DNA marker

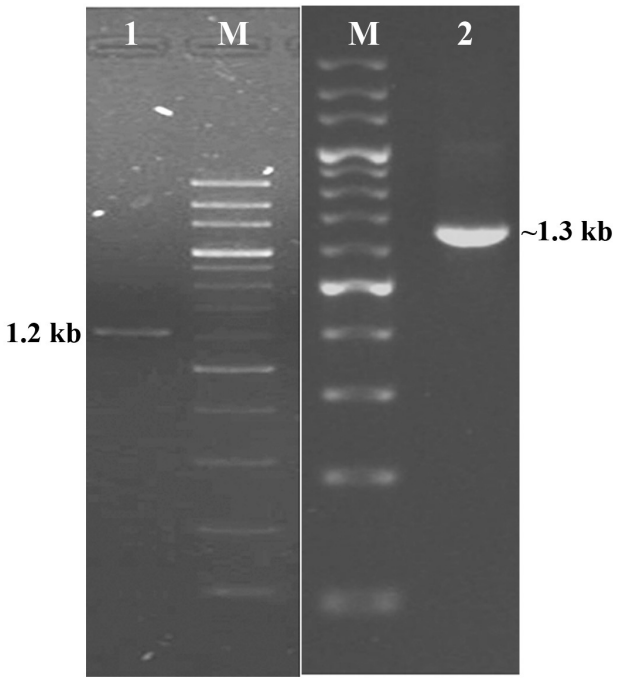


Figure 42. Agarose gel showing the PCR product amplified with *ClPKS11*-FL primers from cDNA (Lane 1) and gDNA (Lane 2). M- 200 bp DNA marker.

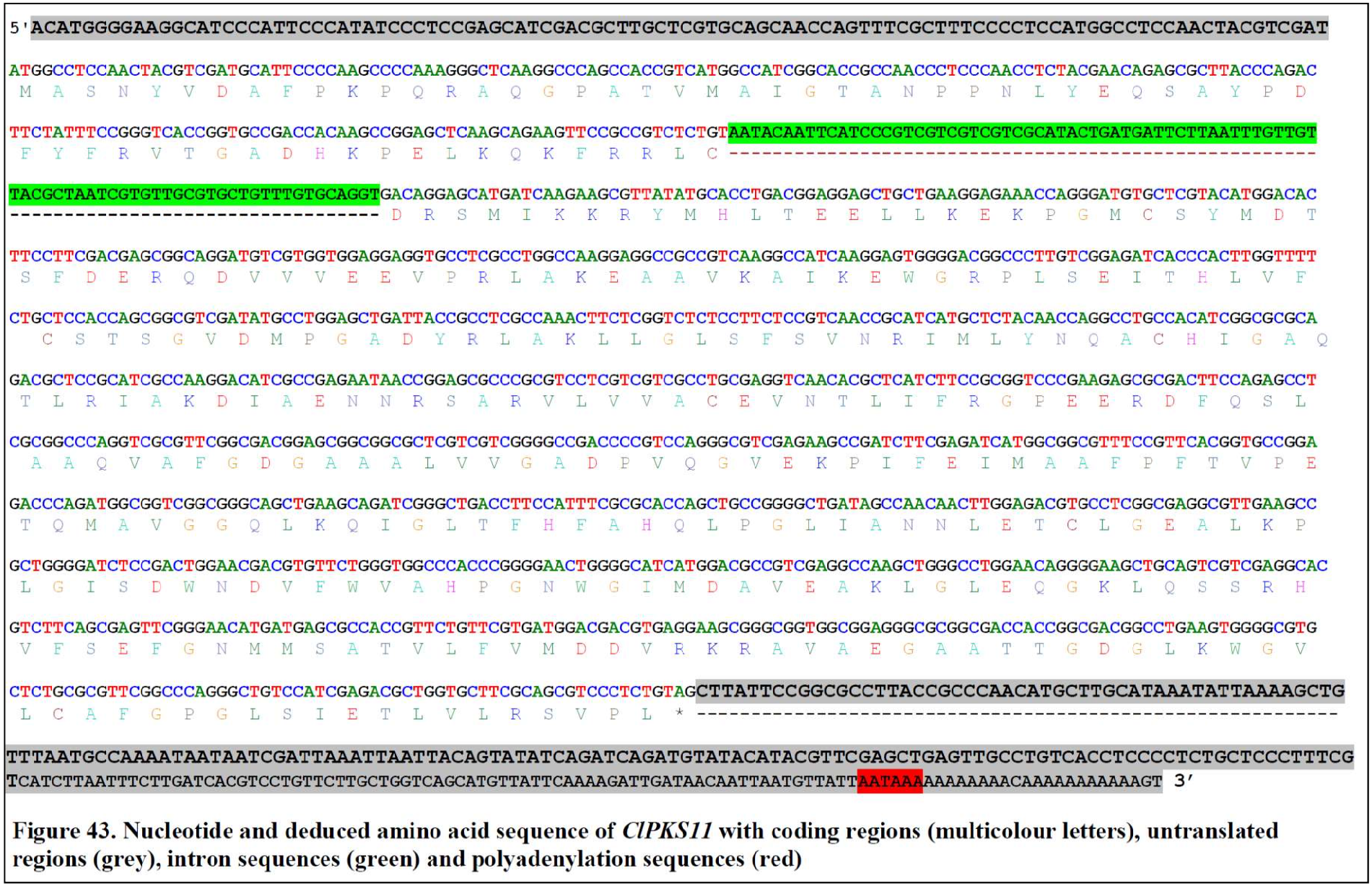


Figure 43. Nucleotide and deduced amino acid sequence of *ClPKS11* with coding regions (multicolour letters), untranslated regions (grey), intron sequences (green) and polyadenylation sequences (red)

UTRScan. Genomic DNA sequence of *ClPKS11* when analyzed using Genscan indicated the presence of an intron of 91 bp. ProtParam predicted molecular mass of 42.9 kDa and a pI of 6.11 for ClPKS11. The amino acid sequence of ClPKS11 deduced using EMBOSS TRANSEQ when compared with NCBI database using BLASTp shared 80% identity with CURS2 like polypeptide of *Musa accuminata*, 72% identity with CURS3 of *C. longa*, 70% identity with CURS1 of *C. longa*, 69% identity with CURS2 of *C. longa* and 61% identity with DCS of *C. longa*. Multiple sequence alignment of ClPKS11 with other PKSs revealed that the catalytic triad residues (Cys164, His303 and Asn336), identical six residue loops (Thr132-Met137), *cis*-peptide bond between Met137 and Pro138 and gatekeepers (Phe215 and Phe265) to block the lower portion of the opening between the CoA-binding tunnel and the active-site cavity were conserved in the ClPKS11 (**Fig. 44**). However, amino acid differences were observed in substrate binding pocket, cyclization pocket, and geometry shapers surrounding the active site (**Table 12**) when analyzed in SBSPKS web server. The data showed that DCS, CURS1, CURS2, CURS3 and ClPKS10 have similar residues lining the substrate binding pocket and CURS1, CURS2, CURS3 and ClPKS10 share similar cyclization pocket. DCS is having single amino acid change in substrate binding pocket compared with all the three isoforms of curcumin synthases.

Table 12. Diversity of amino acid residues in the cyclization pocket, substrate binding pocket and geometry shapers of PKSs

Protein name	Plant	Cyclization pocket	Substrate binding pocket	Geometry shapers
DCS	*C. longa*	AMVGP	SETSQ	PGGLDGPGGGGGG
CURS1	*C. longa*	IMVGP	SETSQ	PAGGDGPGNGGGG
CURS2	*C. longa*	IMVGP	SETSQ	PAGGDGPGNGGGG
CURS3	*C. longa*	IMVGP	SETSQ	PAGGDGPGNGGGG
ClPKS9	*C. longa*	TMIGP	SETTS	PGGLDGPGGGGGA
ClPKS10	*C. longa*	IMVGP	SETSQ	PAGGDGPGNGGGA
CHS	*Medicago sativa* L.	TMIGP	SETTS	PGGLDGPGGGGGA
CUS	*Oryza sativa* L.	NVLMP	SETYS	PGGLDRPGSGGGG
ClPKS11	*C. longa*	TMVGP	SENIM	PAGADGPGNGGGG

Notably, in ClPKS11, residues lining both substrate binding pocket and cyclization pocket were different from other type III PKSs reported from *C. longa*. In ClPKS11, the amino acid residues, Thr132, Ser133, Gly256 and Phe265 were conserved, however Thr194, Thr197 and Ser338 were specifically replaced by Asp, Iso and Met respectively in the corresponding positions of ClPKS11. The 'GFGPG' loop is a region of plant PKSs that forms the cyclization scaffold (Austin and Noel, 2003). In *ClPKS11*, it is replaced by 'AFGPG' (**Fig. 44**). The deduced amino acid sequence of ClPKS11 analyzed from three

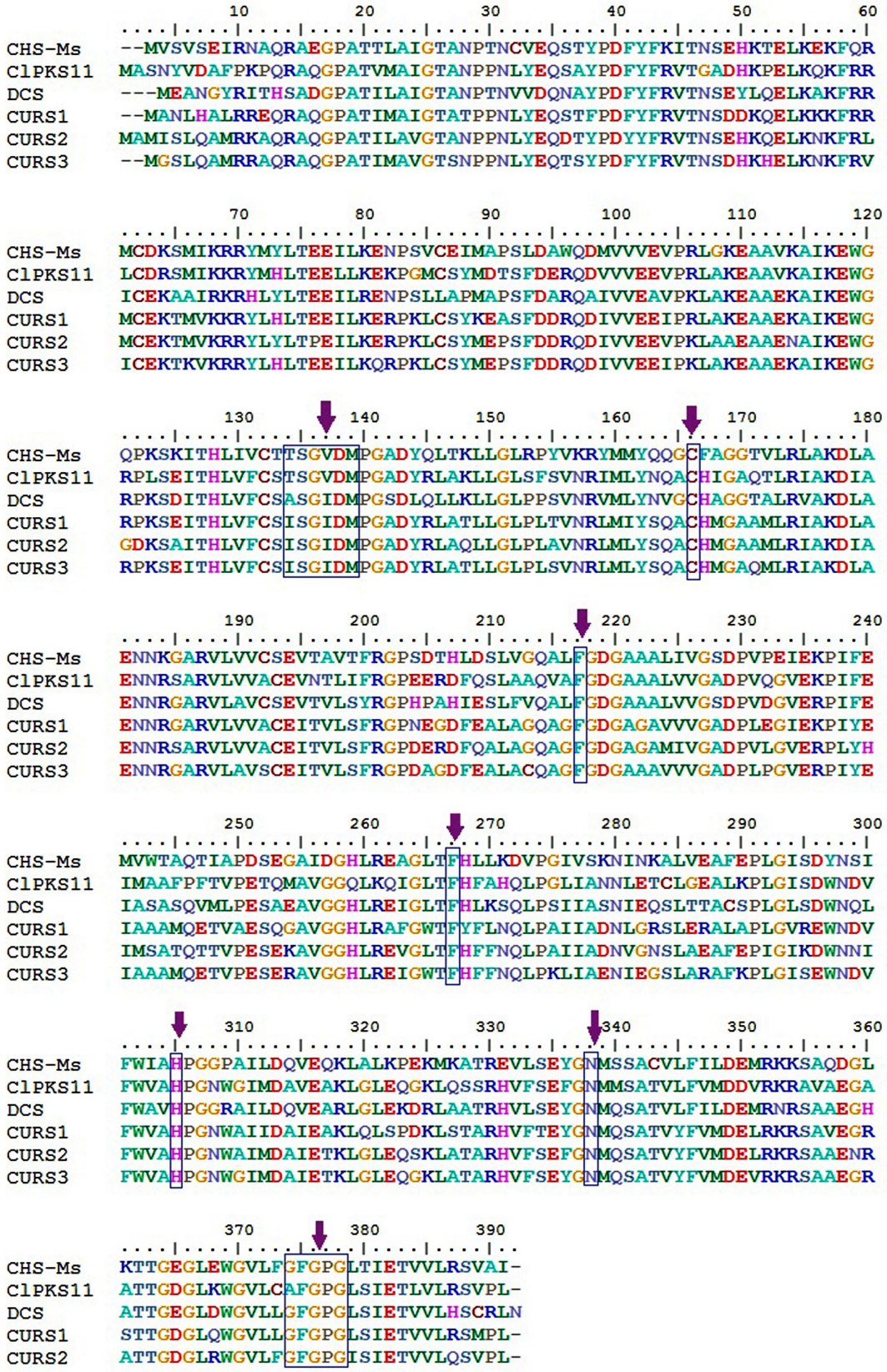


Figure 44. Multiple sequence alignment of ClPKS11 with other type III PKSs. Amino acid residues corresponding to six residue loop (Thr132-Met137), Catalytic triad (Cys164, His303, Asn336) Gate keepers (Phe215, Phe265) and GFGPG loop of CHS-Ms (*Medicago sativa*) are compared with PKSs of *C. longa* and highlighted in boxes.

high curcumin turmeric genotypes [IISR Prathibha (8.0%), Megha turmeric (5.5%), RH17 (10.2%)], three low curcumin genotypes [Collection no. 19 (2.1%), Collection no. 200 (2.1%), Collection no. 449 (2.3%)] and three zero *Curcuma* spp., *C. aromatica, C. amada* and *C. caesia* revealed that the coding sequence of *ClPKS11* is highly conserved irrespective of the variations in curcumin content in these varieties. Phylogenetic analysis showed that ClPKS11 is located together with type III PKS of *M. acuminata* in a cluster containing non-CHS enzymes including three isoforms of curcumin synthase and diketide synthase (**Fig. 45**). ClPKS11 was found to be localized in cytoplasmic matrix exclusive of any transmembrane domain and signal peptide as predicted by SLP-local, SignalP and TMPred server. The secondary structure of ClPKS11 predicted using SOPMA showed that it comprises approximately 41.69% alpha helix, 18.67% extended strand, 9.97% beta turn and 29.67% random coil (**Fig. 46**).

4.9.2 Protein-ligand interaction

Three dimensional (3D) structure of ClPKS11 constructed using Modeller (**Fig. 47**) was confirmed using Ramachandran Plot statistics. The ClPKS11 protein model exhibited a good fit with the reference geometry with 100% of non-glycine and non-proline residues suggesting that ClPKS11 protein model represents a valid stereochemical conformation (**Fig. 48**). The cavity volume of ClPKS11 was (1690 $Å^3$), which is similar to that of CURS1 (1805 $Å^3$) as predicted using CASTP. Molecular docking using Molegro Virtual Docker showed that all the substrates namely, feruloyl CoA, *p*-coumaroyl CoA, feruloyldiketide CoA and *p*-coumaroyldiketide CoA docked satisfactorily to the enzyme active sites of both ClPKS11 and CURS1 with good docking scores. Moldock score, Rerank score, Interaction H-Bond, Docking score and amino acid residues involved in protein–ligand interaction is detailed in **Table 13** and **Fig. 49**. All the substrates showed higher binding affinity to the active site of ClPKS11 than with CURS1. ARG64 of ClPKS11 showed interaction with all the four substrates whereas ASN308 interacted with coumaroyl CoA, coumaroyl diketide CoA and feruloyl diketide CoA and GLY310 interacted with coumaroyl CoA, feruloyl CoA and feruloyl diketide CoA (**Table 14**). Among the substrates, feruloyl CoA exhibited the lowest docking score and its binding with the active site of ClPKS11 is depicted in **Fig. 50**.

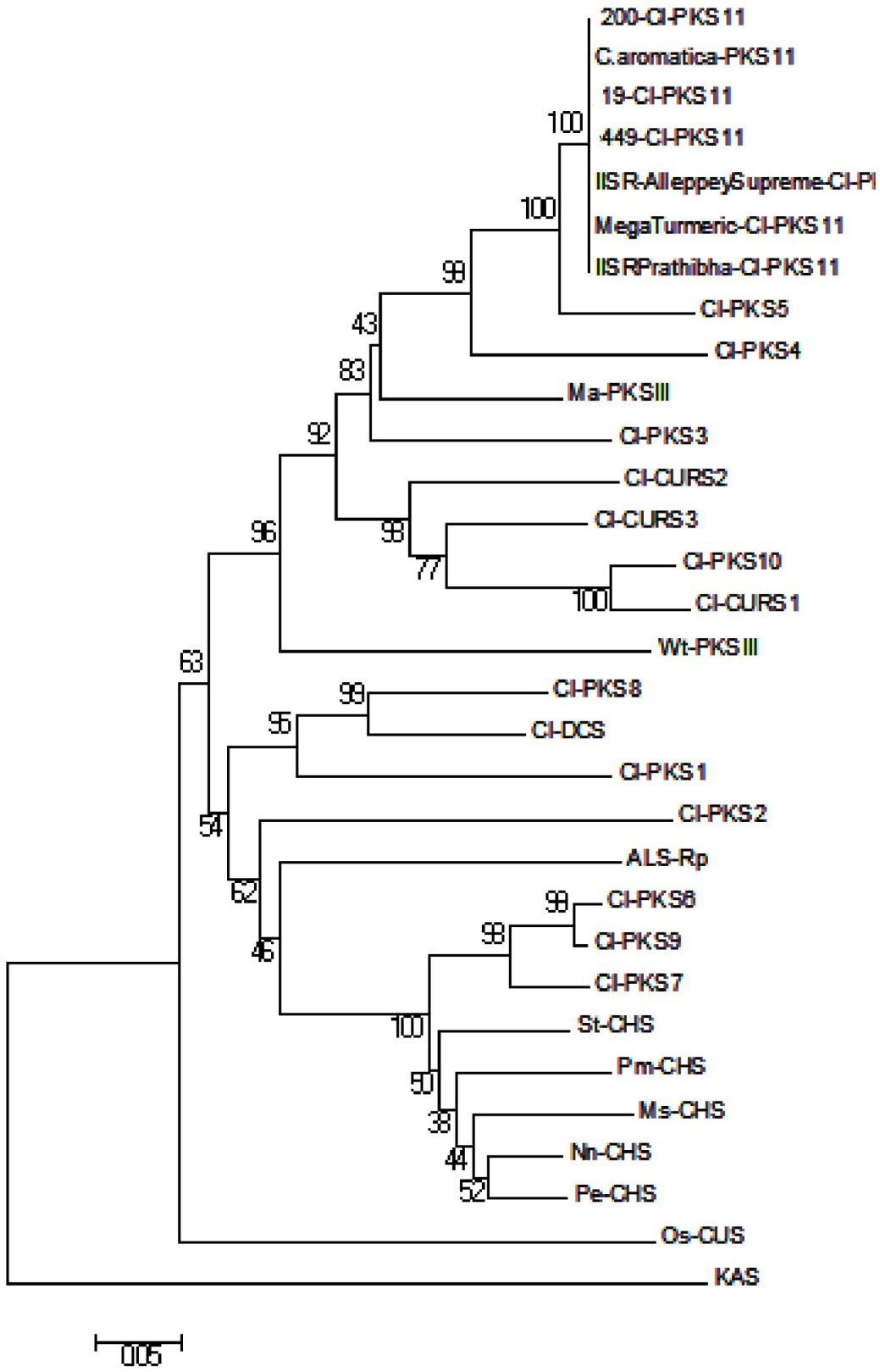


Figure 45. Phylogenetic tree constructed with ClPKS11 and other type III PKSs

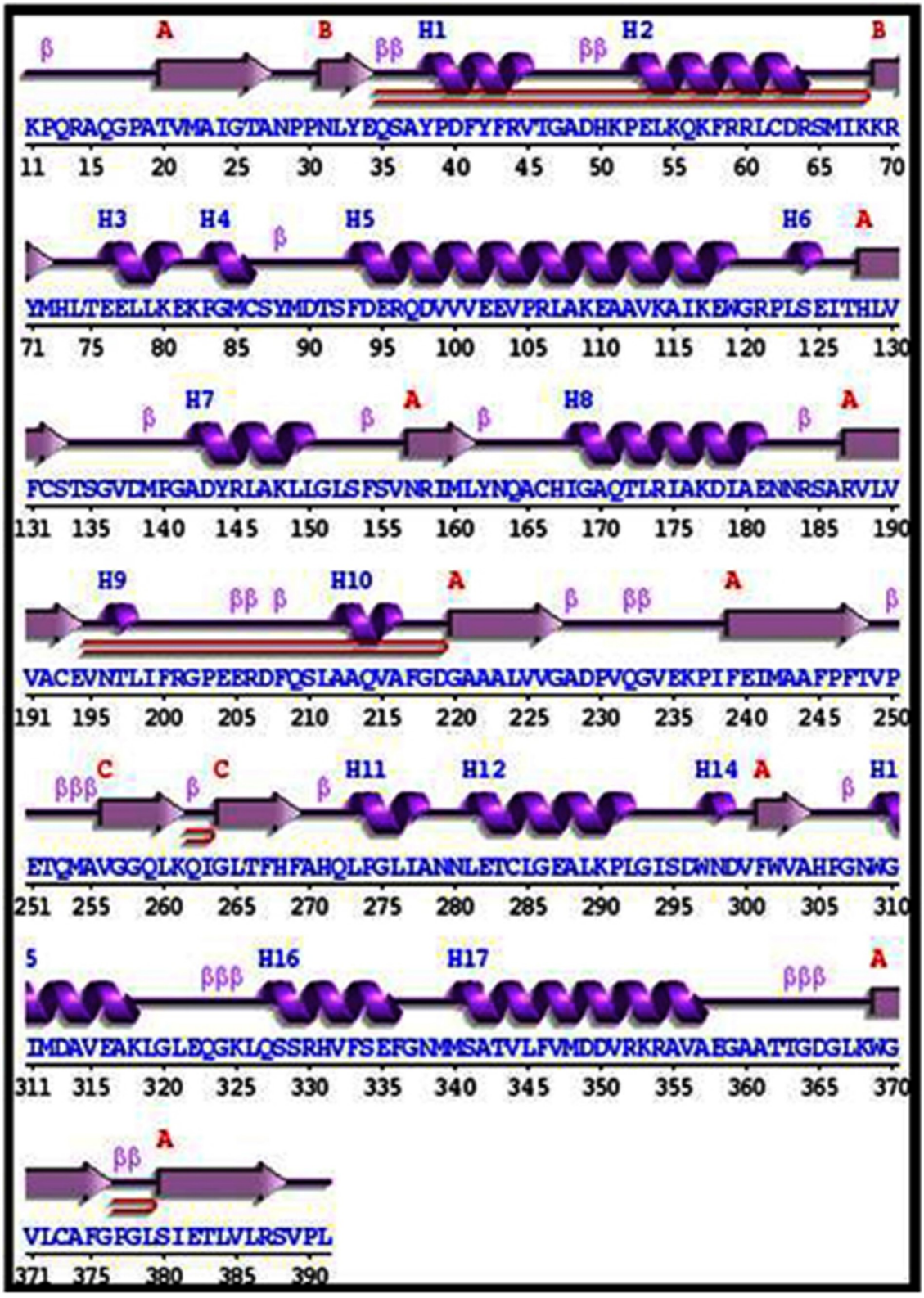


Figure 46. Secondary structure of ClPKS11

Figure 47. Predicted three dimensional structure of ClPKS11 with helices in blue, strands in pink and coils in yellow colour

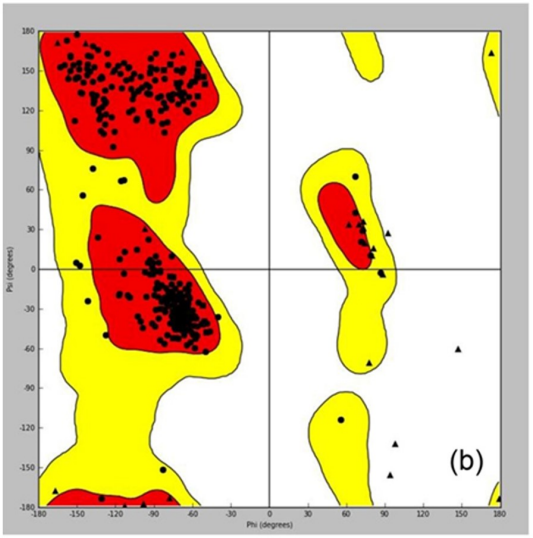


Figure 48. Ramachandran plot analysis of ClPKS11

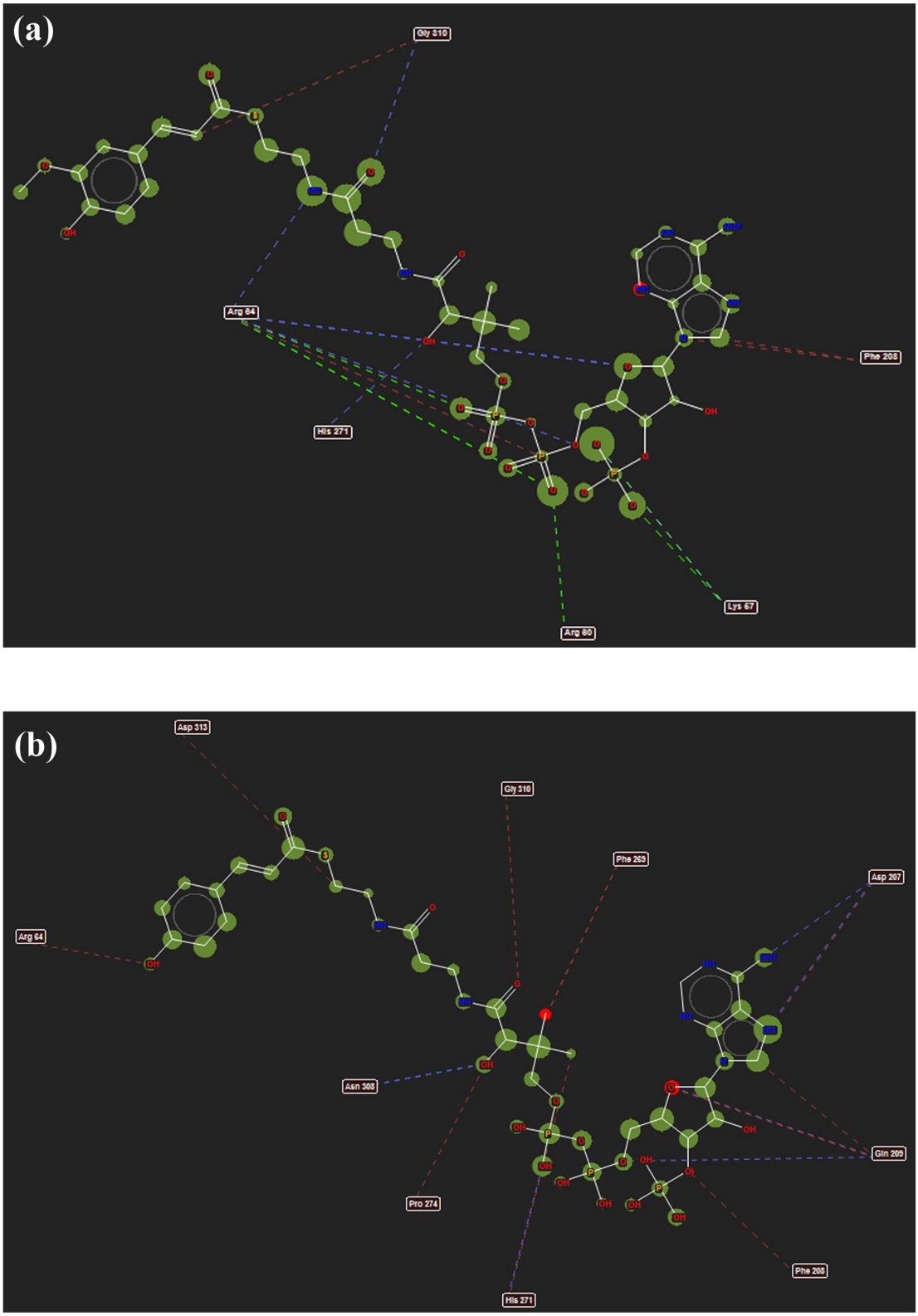


Figure 49. Interaction of substrates (a) feruloyl CoA and (b) coumaroyl CoA with the amino acids of ClPKS11

Fig. 49 continued.....

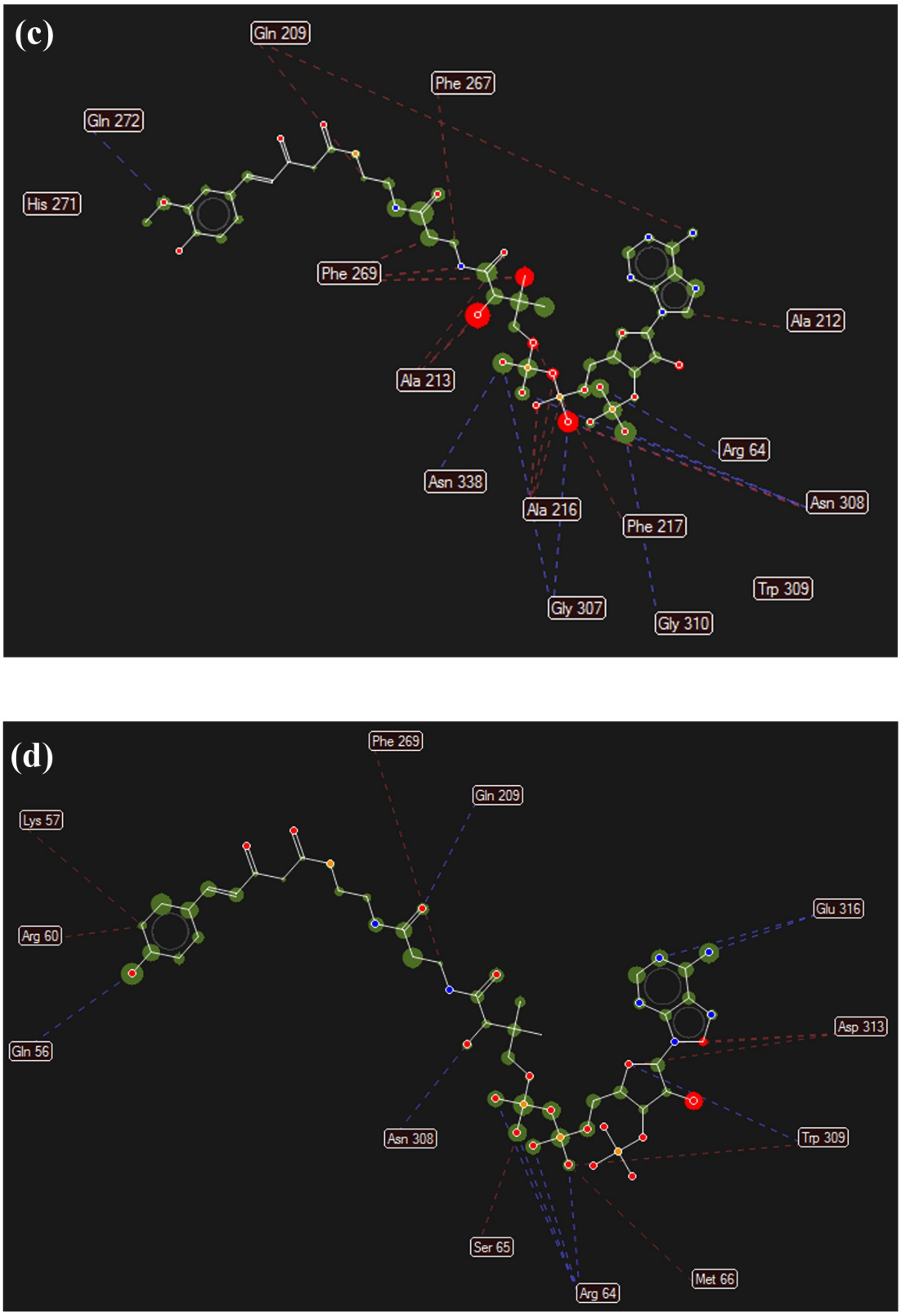


Figure 49. Interaction of substrates .....continued
(c) feruloyldiketide CoA and (d) coumaroyldiketide CoA) with the amino acids of ClPKS11

Table 13. Characteristics of substrates docked with ClPKS11 and CURS1 of *C. longa*

	MolDock Score	Rerank Score	Interaction	HBond	Docking Score
ClPKS11					
Feruloyl CoA	-193.87	-124.138	-226.305	-7.10026	-160.59
Feruloyl diketideCoA	-184.526	-67.2757	-224.661	-5.02847	-169.486
p-coumaroyl diketideCoA	-180.714	44.8108	-226.803	-12.4969	-173.882
p-coumaroyl-CoA	-165.909	-83.3193	-204.388	-5.20376	-163.547
CURS1					
Feruloyl CoA	-165.159	-92.7679	-197.665	-15.5736	-177.335
p-coumaroyl diketideCoA	-164.474	-45.3444	-211.31	-6.35237	-175.48
Feruloyl diketideCoA	-164.365	-68.0237	-194.863	-8.8641	-172.039
p-coumaroyl-CoA	-144.494	-38.7768	-193.034	-6.06322	-157.036

Table 14. Details of amino acid residues involved in ClPKS11-ligand interactions

Ligand name	Hydrogen bond interaction	Steric interaction
Feruloyl CoA	ARG 64, LYS 67, ARG 80, HIS 271,GLY310	ARG 64, PHE 208, GLY 310
Feruloyl diketide CoA	ARG 64, GLN 272, GLY 307, ASN 308, GLY 310, ASN 338	GLN 209, ALA 212, ALA 213, ALA 216, PHE 217, PHE 267, PHE 269, ASN 308
Coumaroyl CoA	ASP 207, GLN 209, HIS 271, ASN 308	ARG 64, ASP 207, PHE 208, GLN 209, HIS 271, PRO 274, GLY 310, ASP 313
Coumaroyl diketide CoA	GLN 56, ARG 64, GLN 209, ASN 308, TRP 309, GLU 316	LYS 57, ARG 60, SER 65, MET 66, PHE 269, TRP 309, ASP 313

4.10 Intron length polymorphism and pre-mRNA analysis of *ClPKS11*

The variation in the length of the intron of *ClPKS11* was detected by sequencing the PCR product amplified with *ClPKS11-FL* primers (Appendix 1). The length of *ClPKS11* intron varied from 88 bp to 95 bp within the genotypes of *C. longa*. Among the six genotypes of *C. longa*, the intron length varied from 88 bp in Collection no. 449 and Megha turmeric to 91 bp in Collection no. 200, IISR Prathibha and Alleppey supreme, and 95 bp in Collection no.19. The intron lengths were 99 bp and 101 bp in *C. aromatica* and *C. caesia* respectively. Megha turmeric was discriminated from other released varieties, IISR Prathibha and IISR Aleppey supreme by three bp insertions. IISR Prathibha and IISR Aleppey supreme were differentiated by one base substitution (C/T) (**Fig. 51**). Further, to determine whether the expression of *ClPKS11* is influenced by post-transcriptional

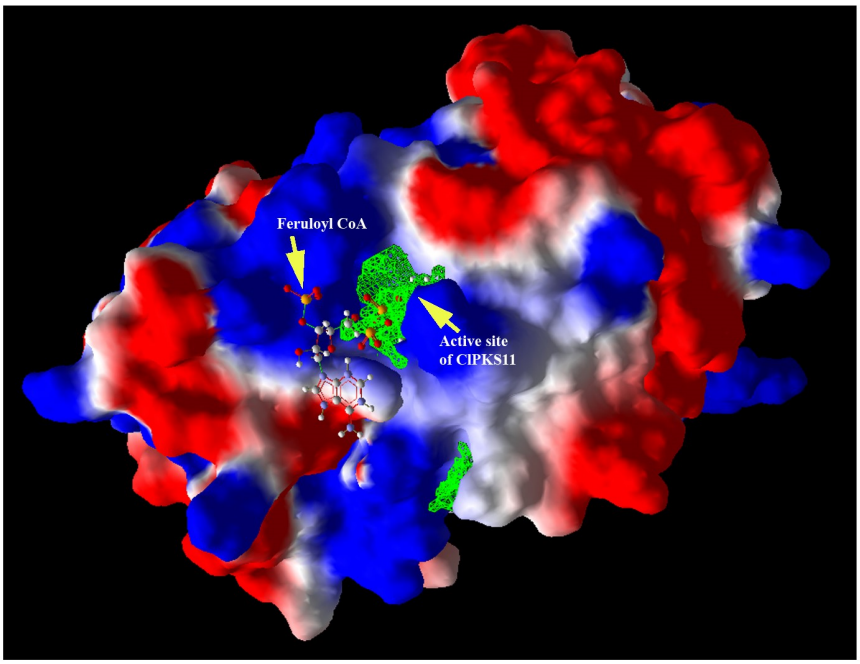


Figure 50. Binding of feruloyl CoA to the active site of ClPKS11

```
                        10         20         30         40         50         60
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
19              1   --------AAT ACAATTCATC CCGTCGTCGT CGTCGTCGCA TAGT---GTT TGATTGATGA  50
200             1   ACCTGCACAA ATAGCACGCA ACACGATTAG CGTAACAACA AATTAA---- -GAATCATCA  55
449             1   --------AAT ACAATTCATC CCGTCGTCGT CGTCG---CA TACT------ -----GATGA  39
Prathibha       1   --------AAC ACAATTCATC CCGTCGTCGT CGTCGTCGCA TACT------ -----GATGA  42
Megaturmeric    1   --------AAT ACAATTCATC CCGTCGTCGT CGTCG---CA TACT------ -----GATGA  39
Aleppey supreme 1   --------AAT ACAATTCATC CCGTCGTCGT CGTCGTCGCA TACT------ -----GATGA  42
C.aromatica     1   ACCTGCACAA ACAGCACGCG ACACGATTAG CATAACAACA AATTAA---- -GAATCATCA  55
C.caesia        1   --------AAT ACATTTCATC CCGTCGTCGT CGTCGTCGCA TACTACTGTT TGATTGATGA  53

                        70         80         90        100        110
                ....|....| ....|....| ....|....| ....|....| ....|....| .
19              51  TTCT--T---  -TGTTGTTAT GCTAATCGTG TTGCGTGCTA TTTGTGCAGG T  95
200             55  ---------GT ATGCGTCGAC GACGACGACG GGATGAATTG TATT------ -  91
449             40  TTCT--TAAT TTGTTGTTAC GCTAATCGTG TTGCGTGCTG TTTGTGCAGG T  88
Prathibha       43  TTCT--TAAT TTGTTGTTAC GCTAATCGTG TTGCGTGCTA TTTGTGCAGG T  91
Megaturmeric    40  TTCT--TAAT TTGTTGTTAC GCTAATCGTG TTGCGTGCTG TTTGTGCAGG T  88
Aleppey supreme 43  TTCT--TAAT TTGTTGTTAC GCTAATCGTG TTGCGTGCTA TTTGTGCAGG T  91
C.aromatica     56  ATCAAACAGT ATTATGCGAC GACGACGAAG GGATGAATTG GATT------ -  99
C.caesia        54  TTCT--TAAT TTGTTGTTTC GCTAATCGTG CTGCGTGCTA TTTGTGCAGG -  101
```

Figure 51. Alignment of *ClPKS11* intron sequences from *Curcuma* spp.

silencing, RT-PCR was carried out with pre-mRNA primer, designed from exon1-intron junction and mature mRNA primer designed from exon 1-exon 2 region. The amplification was observed with mature mRNA primers; *ClPKS11* exon1 and *ClPKS11* exon 2 (Appendix 1) from both IISR Prathibha (high curcumin genotype) and Collection no. 200 (low curcumin genotype), with higher intensity in IISR Prathibha whereas, no amplification was observed with pre-mRNA primers; *ClPKS11* exon1 and *ClPKS11* intron (Appendix 1) from both the samples (**Fig. 52**) in agarose gel electrophoresis. However, similar levels of pre-mRNA were detected from both the turmeric genotypes through qPCR.

4.11 Identification of upstream sequence of *ClPKS11*

The expression of many structural genes is controlled by the binding of transcription factors to the sequences upstream of coding region. To understand the probable transcription factors regulating ClPKS11, genomic DNA-inverse PCR method was used to amplify the sequence upstream of *ClPKS11*. The restriction sites within the ORF of *ClPKS11* were identified using NEBcutter and represented in **Fig. 53**. The restriction enzyme, *Pst*I was selected and genomic DNA was digested followed by circularization of restricted fragments. Inverse PCR reaction using outward designed primers; PMTR-IPa (Appendix 1) followed by nested PCR with PMTR-IPb primers (Appendix 1) amplified 0.8 kb DNA fragment (**Fig. 54a**). The PCR product was cloned and sequenced. Sequence analysis involving the removal of vector sequences and known sequence of *ClPKS11*, resulted in the identification of a 495 bp sequence upstream of the start codon of *ClPKS11* (**Fig. 54b**). The sequence was further confirmed by end to end PCR with primers; PMTR (Appendix 1) designed from the coding region of *ClPKS11* (**Fig. 54c**). Transcription start site (TSS) was predicted to be located at 201 bp and TATA box with maximum score was predicted to be located at 168 bp. The promoter elements; TATA box (-33 bp away from the TSS) and CAAT box were also predicted from the upstream region of ClPKS11. Many transcription factor families like MYB, bHLH, WRKY are reported as the regulators of secondary metabolite biosynthesis (Zhai et al., 2015; Huang et al., 2016; Amato et al., 2016; Xiang et al., 2015). Twenty two *cis*–regulatory elements were predicted using PlantCARE database from both strands of *ClPKS11*, which involves MYB and WRKY binding sites also (**Table 15**). PlantRegMap identified 44 putative MYB binding sites in the upstream sequence of *ClPKS11* (**Table 16**).

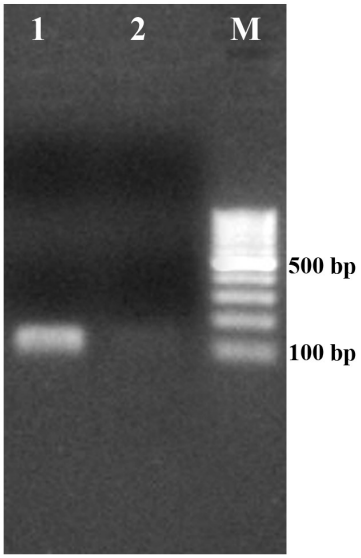


Figure 52. Agarose gel showing PCR amplification of mature mRNA of *ClPKS11* with exon1-exon2 primers in cDNA of IISR Prathibha (Lane 1) and Collection no. 200 (Lane 2). M-100 bp DNA ladder.

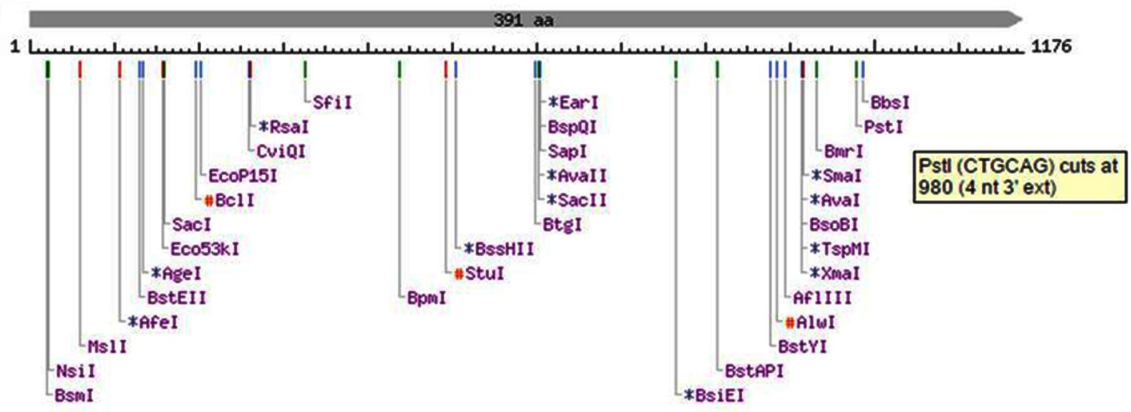


Figure 53. Identification of restriction enzyme sites in the ORF of *ClPKS11*

Table 15 a. List of putative *cis*-acting regulating elements in the sense (positive) strand of upstream sequence of *ClPKS11*

Site Name	Organism	Position	Sequence	Function
TC-rich repeats	*Nicotiana tabacum*	4	ATTTTCTTCA	*cis*-acting element involved in defense and stress responsiveness
TCT-motif	*Arabidopsis thaliana*	18	TCTTAC	part of a light responsive element
MBS	*Zea mays*	25	CGGTCA	MYB Binding Site
TGACG-motif	*Hordeum vulgare*	52	TGACG	*cis*-acting regulatory element involved in the MeJA-responsiveness
TATA-box	*Glycine max*	83	TAATA	core promoter element around -30 of transcription start
TATA-box	*Brassica napus*	85	ATATAT	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	86	TATA	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	88	TATA	core promoter element around -30 of transcription start
CAAT-box	*Brassica rapa*	118	CAAAT	common *cis*-acting element in promoter and enhancer regions
TATA-box	*Arabidopsis thaliana*	135	TATA	core promoter element around -30 of transcription start
TATA-box	*Brassica napus*	166	ATATAT	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	167	TATA	core promoter element around -30 of transcription start
TATA-box	*Brassica napus*	168	ATATAT	core promoter element around -30 of transcription start
TATA-box	*Arabidopsis thaliana*	169	TATA	core promoter element around -30 of transcription start
Unnamed__2	*Triticum aestivum*	258	CCACGTCACCG	single-strand DNA-binding proteins site (ssDBP-1 and -2)
G-box	*Zea mays*	259	CACGTC	*cis*-acting regulatory element involved in light responsiveness
CGTCA-motif	*Hordeum vulgare*	261	CGTCA	*cis*-acting regulatory element involved in the MeJA-responsiveness
CCAAT-box	*Hordeum vulgare*	264	CAACGG	MYBHv1 binding site
MBS	*Zea mays*	267	CGGTCA	MYB Binding Site
Box-W1	*Petroselinum crispum*	278	TTGACC	fungal elicitor responsive element
W box	*Arabidopsis thaliana*	278	TTGACC	
Unnamed__4	*Petroselinum hortense*	321	CTCC	
CAAT-box	*Glycine max*	331	CAATT	common *cis*-acting element in promoter and enhancer regions
HSE	*Brassica oleracea*	336	AGAAAATTCG	*cis*-acting element involved in heat stress responsiveness
CAAT-box	*Arabidopsis thaliana*	359	gGCAAT	common *cis*-acting element in promoter and enhancer regions
CAAT-box	*Hordeum vulgare*	361	CAAT	common *cis*-acting element in promoter and enhancer regions
TATA-box	*Arabidopsis thaliana*	402	taTATAAAtc	core promoter element around -30 of transcription start
ATCT-motif	*Pisum sativum*	408	AATCTAATCC	part of a conserved DNA module involved in light responsiveness
CAAT-box	*Arabidopsis thaliana*	411	CCAAT	common *cis*-acting element in promoter and enhancer regions
ATCC-motif	*Pisum sativum*	412	CAATCCTC	part of a conserved DNA module involved in light responsiveness
CAAT-box	*Hordeum vulgare*	412	CAAT	common *cis*-acting element in promoter and enhancer regions
Unnamed__4	*Petroselinum hortense*	450	CTCC	

Table 15 b. List of putative *cis*-acting regulating elements in the anti-sense (negative) strand of upstream sequence of *ClPKS11*

Site Name	Organism	Position	sequence	Function
GA-motif	*Glycine max*	11	AAGGAAGA	part of a light responsive element
Box-W1	*Petroselinum crispum*	26	TTGACC	fungal elicitor responsive element
W box	*Arabidopsis thaliana*	26	TTGACC	
CGTCA-motif	*Hordeum vulgare*	52	CGTCA	*cis*-acting regulatory element involved in the MeJA-responsiveness
TATA-box	*Arabidopsis thaliana*	134	TATAA	core promoter element around -30 of transcription start
CAAT-box	*Hordeum vulgare*	172	CAAT	common *cis*-acting element in promoter and enhancer regions
MBS	*Arabidopsis thaliana*	191	CAACTG	MYB binding site involved in drought-inducibility
CAAT-box	*Brassica rapa*	213	CAAAT	common *cis*-acting element in promoter and enhancer regions
ABRE	*Arabidopsis thaliana*	257	ACGTGGC	*cis*-acting element involved in the abscisic acid responsiveness
TGA-box	*Glycine max*	257	TGACGTGGC	part of an auxin-responsive element
G-box	*Brassica napus*	258	TGACGTGG	*cis*-acting regulatory element involved in light responsiveness
O2-site	*Zea mays*	258	GTTGACGTGA	*cis*-acting regulatory element involved in zein metabolism regulation
Unnamed__1	*Zea mays*	258	CGTGG	
Unnamed__3	*Zea mays*	258	CGTGG	
TGACG-motif	*Hordeum vulgare*	261	TGACG	*cis*-acting regulatory element involved in the MeJA-responsiveness
Box-W1	*Petroselinum crispum*	268	TTGACC	fungal elicitor responsive element
W box	*Arabidopsis thaliana*	268	TTGACC	
L-box	*Petroselinum crispum*	285	TCTCACCTACC	part of a light responsive element
Unnamed__2	*Zea mays*	301	CCCCGG	
GARE-motif	*Brassica oleracea*	369	TCTGTTG	gibberellin-responsive element
C-repeat/DRE	*Arabidopsis thaliana*	383	TGGCCGAC	regulatory element involved in cold- and dehydration-responsiveness
TATA-box	*Helianthus annuus*	400	TATACA	core promoter element around -30 of transcription start
TATA-box	*Glycine max*	404	TAATA	core promoter element around -30 of transcription start
Box 4	*Petroselinum crispum*	405	ATTAAT	part of a conserved DNA module involved in light responsiveness

Table 16. List of probable MYBs binding to the upstream sequence of *ClPKS11*

S. No.	MYB	Strand	Position	p-value	Matched sequence
1	MYB51	-	284-304	1.90E-06	CCGGTTCGATTTCACCTACCA
2	MYB13	+	280-294	3.11E-06	GACCTGGTAGGTGAA
3	MYB17	+	280-294	3.74E-06	GACCTGGTAGGTGAA
4	MYB3R-5	-	264-278	4.23E-06	AAGAGTTGACCGTTG
5	MYB84	+	281-293	4.78E-06	ACCTGGTAGGTGA
6	MYB61	-	280-294	5.34E-06	TTCACCTACCAGGTC

7	MYB3R1	-	263-277	5.78E-06	AGAGTTGACCGTTGA
8	MYB92	+	285-294	6.18E-06	GGTAGGTGAA
9	MYB15	-	284-304	6.27E-06	CCGGTTCGATTTCACCTACCA
10	MYB3R-4	-	263-277	6.79E-06	AGAGTTGACCGTTGA
11	MYB10	+	279-294	7.93E-06	TGACCTGGTAGGTGAA
12	MYB12	-	284-293	1.32E-05	TCACCTACCA
13	MYB99	-	283-295	1.34E-05	TTTCACCTACCAG
14	MYB3	+	284-294	1.37E-05	TGGTAGGTGAA
15	MYB39	+	282-295	1.80E-05	CCTGGTAGGTGAAA
16	MYB40	-	282-294	2.05E-05	TTCACCTACCAGG
17	MYB43	+	283-295	2.06E-05	CTGGTAGGTGAAA
18	MYB63	-	227-247	2.40E-05	CGATCAACTCAACTACCCTGA
19	MYB58	+	282-296	2.40E-05	CCTGGTAGGTGAAAT
20	MYB46	+	284-293	2.44E-05	TGGTAGGTGA
21	MYB41	-	284-298	2.54E-05	CGATTTCACCTACCA
22	MYB23	+	229-239	2.65E-05	AGGGTAGTTGA
23	MYB111	+	284-293	2.96E-05	TGGTAGGTGA
24	MYB4	+	280-294	3.06E-05	GACCTGGTAGGTGAA
25	MYB99	-	229-241	3.37E-05	ACTCAACTACCCT
26	MYB55	-	282-295	3.49E-05	TTTCACCTACCAGG
27	MYB67	+	274-294	3.81E-05	CTCTTTGACCTGGTAGGTGAA
28	MYB60	-	280-298	4.01E-05	CGATTTCACCTACCAGGTC
29	MYB74	-	282-295	4.15E-05	TTTCACCTACCAGG
30	MYB15	-	230-250	4.20E-05	GGTCGATCAACTCAACTACCC
31	MYB13	+	226-240	4.20E-05	CTCAGGGTAGTTGAG
32	MYB60	-	226-244	4.26E-05	TCAACTCAACTACCCTGAG
33	MYB84	+	301-313	4.52E-05	CCGGGGTTGGTCT
34	MYB94	+	226-244	4.83E-05	CTCAGGGTAGTTGAGTTGA
35	MYB17	+	226-240	5.27E-05	CTCAGGGTAGTTGAG
36	MYB80/103	+	281-295	5.44E-05	ACCTGGTAGGTGAAA
37	MYB77	-	264-278	5.69E-05	AAGAGTTGACCGTTG
38	MYB107	+	229-246	6.38E-05	AGGGTAGTTGAGTTGATC
39	MYB63	-	281-301	6.44E-05	GTTCGATTTCACCTACCAGGT
40	MYB94	+	280-298	6.51E-05	GACCTGGTAGGTGAAATCG
41	MYB49	-	283-297	6.55E-05	GATTTCACCTACCAG
42	MYB84	+	227-239	7.06E-05	TCAGGGTAGTTGA
43	MYB96	-	318-335	8.68E-05	AATTGCAGATGGGAGTGG
44	MYB58	+	228-242	9.43E-05	CAGGGTAGTTGAGTT
45	MYB83	+	283-293	9.60E-05	CTGGTAGGTGA

4.12 Proteomics approach to identify differentially expressed proteins

Phenol extraction based method was used to extract protein from rhizomes of IISR Prathibha (high curcumin) and Collection no. 449 (low curcumin). The concentration of

protein, analyzed using Bradford's assay with BSA (Bovine Serum Albumin) as standard was in the range of 1.8±0.6 mg/g of rhizome. The quality of protein was analyzed in SDS PAGE and it showed the presence of intact bands (**Fig. 55**). To identify the proteins differentially expressed in turmeric genotypes with contrasting curcumin content, 150 µg of protein from each sample was subjected to isoelectric focussing on IPG strips pH 4-7 (13 cm) and then separated on 12% SDS PAGE. The gels were stained and 42 protein spots were identified as differentially expressed in turmeric genotypes with contrasting curcumin content (**Fig. 56**).

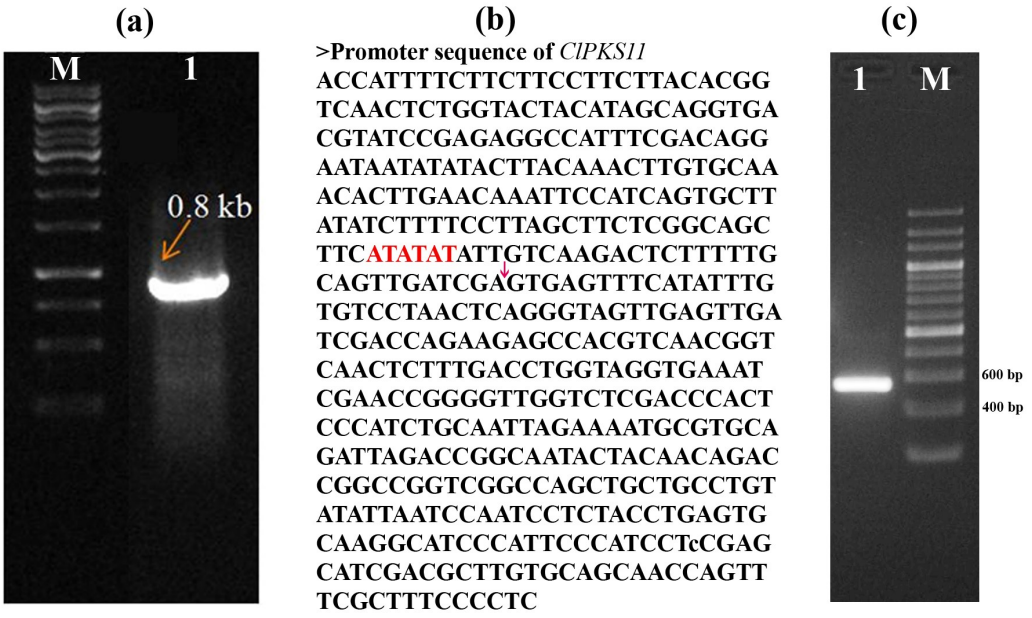


Figure 54. (a) Agarose gel showing the PCR product of gDNA -inverse PCR with PMTR-IPb primers (Lane 1). M-1 kb DNA marker
(b) Upstream sequence of ClPKS11. TSS at 201 bp and TATA box with maximum score (168 bp) are highlighted
(c) Confirmation of promoter sequence using PCR amplification with PMTR primers (Lane 1). M-200 bp DNA ladder

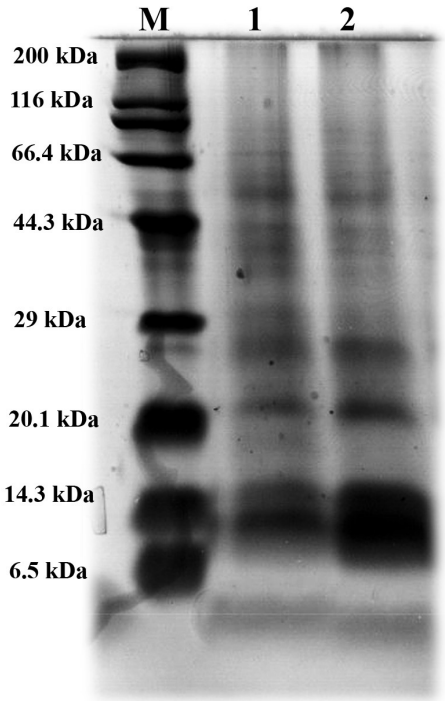


Figure 55. SDS-PAGE (12%) analysis of protein extracted from rhizomes of (1) IISR Prathibha and (2) Collection no. 449. M-Protein molecular weight marker

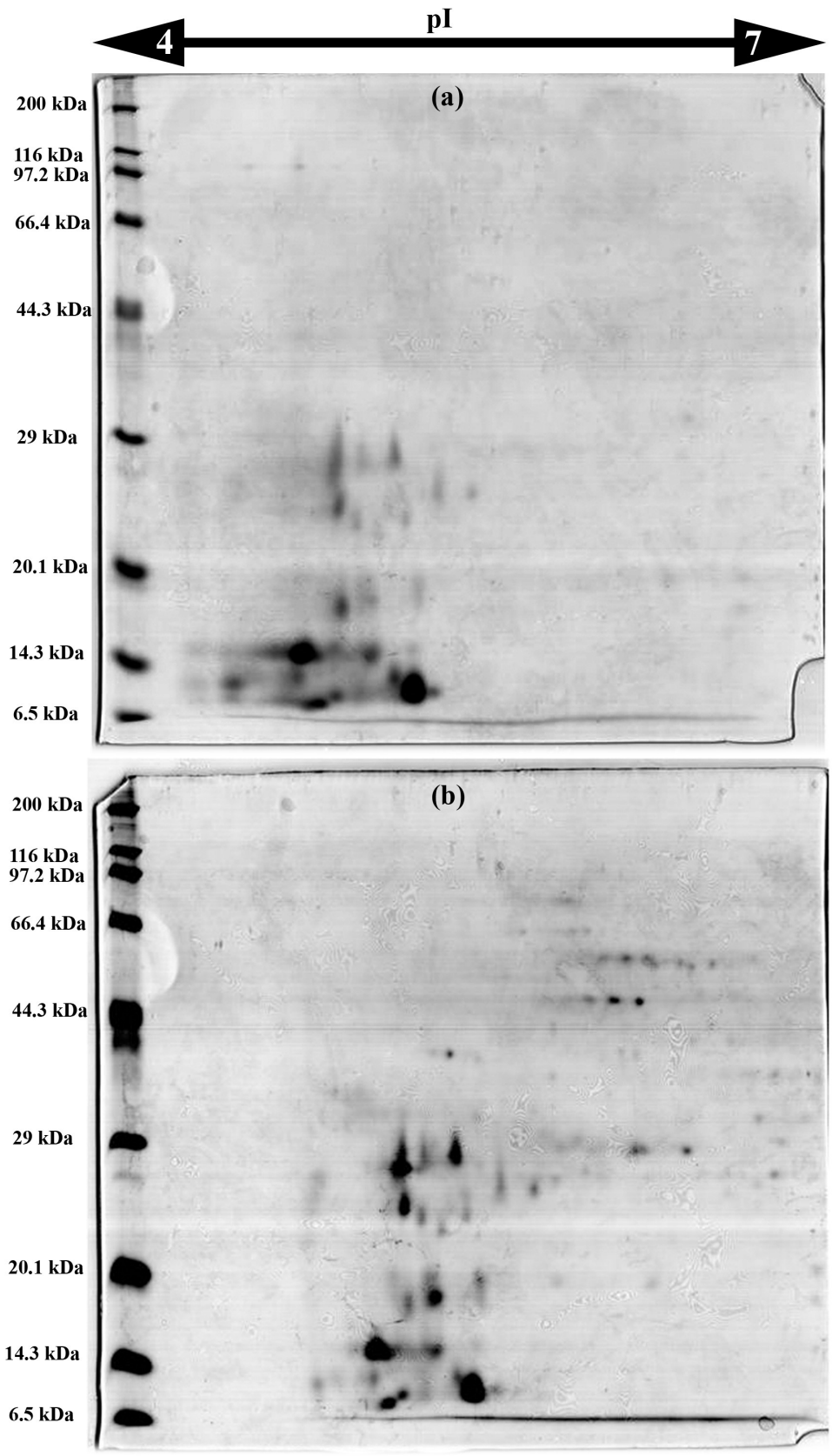


Figure 56. 2DE of protein extracted from rhizomes of IISR Prathibha (a) and Collection no. 449 (b) separated by IEF in 13 cm long IPG strip (pH 4-7) followed by SDS-PAGE (12%). Protein detection by coomassie brilliant blue R250 staining.

Discussion...

Discussion

Curcuma longa (L.) is a medicinal spice and has been valued for therapeutical properties since immemorial. It is commonly known as turmeric and commercially 'turmeric' refers to the dried rhizomes of *Curcuma longa*. There are ample reports establishing the beneficial activities of turmeric for treating various ailments like digestive disorders, cardiac diseases, diabetes, Alzheimer's disease, cancer etc. Most of the therapeutic activities of turmeric are attributed to curcumin which is a polyphenol (Ravindran et al., 2007). Indian turmeric is considered as the best because of its high curcumin content (Sasikumar, 2005). Therapeutic properties of curcumin and rising consumer health consciousness for natural colouring/flavouring substances and herbal/ayurvedic skin care products augment its demand in pharmaceuticals, food and cosmetic industries. India is the largest manufacturer of curcumin, with over 80% of the global market and majority was exported to Europe and North America. The key vendor in the curcumin market includes Synthite Industries, Biomax Life Sciences, Hindustan Mint & Agro Products, SV Agrofood, Arjuna Natural Extracts, Star Hi Herbs, Herboveda India, Helmigs Prima Sejahtera, Tri Rahardja PT, Hebei Food Additive, and Konark Herbals & Health Care. Major cosmetic products using curcumin as ingredient includes Vicco, Himalaya Herbals and Shahnaz Husain. It is reported that global cancer drug and organic cosmetic market has surpassed USD 100 billion and USD 7 billion respectively in 2015 and expected to reach nearly USD 145 billion and USD 13 billion by 2024 (https://www.gminsights.com/industry-analysis/curcumin-market). There are a lot of curcumin supplements in the market in the form of capsules, tablets, soft gels or powder. To enhance the bioavailability of curcumin, several formulations have been prepared which include nanoparticles, liposomes, micelles, and phospholipid complexes. These formulations were reported to have better efficacy in their biological properties compared to free curcumin (Priyadarsini, 2014).

Curcumin content however varies within turmeric accessions (Wang et al., 1999, Jayaprakasha et al., 2002) and with agro-climatic conditions (Anandaraj et al., 2014), which is a major constrain for curcumin industries. Understanding the curcumin biosynthesis, its regulation and the factors altering the curcumin content may aid in maintaining its stability and increase in content across agro-climatic zones. Curcumin is a phenylpropanoid derivative (Roughley and Whiting, 1973; Neema, 2005; Ramirez-Ahmuda et al., 2006). Schroder (1997) hypothesized that type III polyketide synthases

(PKS) are involved in curcumin biosynthesis. Later Katsuyama et al. (2009a and 2009b) identified and characterized four PKSs from turmeric, namely diketide CoA synthase (DCS), curcumin synthase (CURS 1-3) to be involved in curcumin biosynthesis from carboxylic acid CoA ester. The transcription factors like MYB were also suggested to have regulatory roles in curcumin biosynthesis (Koo et al., 2013). Curcuminoids are generally accumulated in the rhizomes of turmeric and hence to study gene expression analysis, we were in need of an efficient protocol for isolation of good quality RNA from turmeric rhizomes as well as other plant parts. Thus our first effort was to optimise a rapid protocol to isolate functional RNA from turmeric tissues suitable for downstream applications like transcriptome sequencing and qRT-PCR.

5.1 Optimization of RNA isolation protocol from turmeric tissues

The extraction of good quality RNA is essential for various downstream applications and the major hindrance in isolating RNA from recalcitrant tissues is the presence of polyphenols and polysaccharides in large quantities (Salzman et al., 1999; Peng et al., 2014; Kansal et al., 2008; Dash, 2013). Turmeric rhizomes are rich in both polysaccharides and polyphenols (Ravindran et al., 2007), thus possessing an obstacle to extract RNA. Polyphenolic compounds oxidize very quickly during RNA extraction and form quinones which bind with RNA (Loomis, 1974) and polysaccharides coprecipitate with the RNA in low ionic concentration buffer (Birtic and Kranner, 2006). In the present study, four protocols reported to isolate RNA from recalcitrant tissues (Chomczynski and Sacchi, 1987; Hou et al., 2011; Chang et al., 1993 and Ghawana et al., 2011) were tried to isolate RNA from leaves and rhizomes of turmeric. All the protocols worked well with leaves but failed in case of rhizomes. Plant RNeasy kit was also tried in rhizomes but the yield was negligible. Hence, optimization of rapid protocol to isolate good quality RNA from turmeric rhizomes was vital in this scenario.

The protocol of Hou et al. (2011) was opted for modification as it was rapid and simple. Moreover an increase in sample: extraction buffer volume from 1: 1.2 to 1: 20 significantly increased the RNA yield. The extraction buffer mentioned in Hou et al. (2011) has Tris-HCl, sodium chloride, SDS and β-mercaptoethanol. In turmeric rhizomes, the addition of sodium chloride decreased the RNA yield as reported in wheat (Singh et al., 2003). But the presence of PVP and EDTA enhanced the yield. So the ingredients of extraction buffer were changed and the concentration of SDS, PVP and β-mercaptoethanol

was optimized. SDS is widely used as a cell disrupting agent in RNA isolation (Ghawana et al., 2011; Hou et al., 2011; Gonzalez et al., 2008) and to inhibit RNase activity (Mendelsohn and Young 1978). A method employing PVP was successful in extracting RNA from leaves and flower buds of *Dendrobium perianth* (Champagne and Kuehnle 2000). β-mercaptoetanol was also used to inhibit RNase activity and to prevent sample oxidation (Gonzalez et al., 2008). The acidic pH was maintained by the addition of acid phenol as previously described to decrease the DNA contamination (Chomczynski and Sacchi 1987). Phenol: chloroform extraction aided in the protein precipitation (Perry and Kelley, 1972). In addition, this step aid in eliminating the plant pigments like curcumin, which is a polyphenol and chlorophyll. Chloroform: isoamylalcol treatment was efficient in removing the carotenoids of brown algae (Greco et al., 2014). In turmeric rhizomes, re-extraction of the aqueous phase along with the sodium acetate and acid phenol: chloroform further decreased the polysaccharide, protein and polyphenol contamination. Sodium acetate (3 M) precipitation was reported to increase the RNA yield by eliminating polysaccharide contamination (Rubio-Pina and Zapata-Perez, 2011; Singh et al., 2003; Suzuki et al., 2004; Wang et al., 2005a). Further, for the enhancement of RNA precipitation, sodium acetate was added along with isopropanol.

The optimized protocol was successfully extended to isolate RNA from fresh as well as RNAlater stored leaves, pseudostem and roots of turmeric and from rhizomes of related species. The good quality of RNA was confirmed using spectrophotometric and electrophoretic analysis. The values of A_{260}/A_{280} analyzing the protein contamination and A_{260}/A_{230} measuring the organic contamination like polyphenol and polysaccharide of total RNA extracted using optimized protocol was in the accepted range, indicating the RNA of good quality and amenable for downstream applications. Analysis using Agilent Bioanalyzer 2100 further confirmed high quality RNA (RIN>7) from different tissues of turmeric isolated by this protocol. A RIN of greater than or equal to seven indicates the RNA is suitable for high stringency applications such as cDNA library construction and next generation sequencing (Schroeder et al., 2006). Total RNA examined in 1.2% denaturing electrophoresis showed distinct rRNA bands with higher or equal intensity of 28s rRNA band than 18SrRNA band, indicating minimal RNA degradation during extraction. Additional rRNA bands were also obtained from the leaf and pseudostem tissues. These bands represent the abundant transcripts as these tissues are photosynthetically active (Masek et al., 2005). Small RNA fraction was enriched from total

RNA and when analyzed in 15% denaturing PAGE noted intact bands of 5.8SrRNA, 5SrRNA, tRNA and microRNAs, ensuring the quality of total RNA. High quality total RNA is a requisite for experimental analysis of microRNAs and other small RNAs (Accerbi et al., 2010).

The functional intactness of isolated mRNAs can be measured by RT-PCR analysis since reverse transcriptase is highly sensitive to impurities (Li et al., 2006; Ghangal et al., 2009). The RNA extracted using optimized protocol was reverse transcribed and actin fragment of ~950 bp was amplified. qRT PCR was also successfully performed with actin primers using cDNA synthesized from total RNA. The expression profile of a novel microRNA was also analyzed to authenticate the functionality of small RNAs enriched from total RNA extracted using optimized protocol. The amplification was observed in rhizome and root tissues but not in leaf and pseudostem tissues, which may be due to tissue specificity (Thatcher et al., 2015). This protocol which was optimized to isolate RNA from turmeric was later reported to be effective in extracting RNA from the roots of *Santalum album,* where many other protocols were futile (Deepa and Yusuf, 2016).

5.2 Construction of full length normalized cDNA library

The lack of genomic sequencing data limited the gene identification and understanding the specific pathway leading to unique secondary metabolites in many non model plants (Tao et al., 2011; Tian et al., 2015; Mazumdar and Chattopadhyay, 2015). Similarly, turmeric also does not have a genomic database. ESTs from rhizomes (6471) and ESTs from leaves (6769) of turmeric were available in the Aromatic Rhizome EST (ArREST) database (http://www.agcol.arizona.edu/cgi-bin/pave/GT/index.cgi#library). Annadurai et al. (2013) developed transcriptomes from three cultivars of turmeric using Illumina GAIIx platform and resulted in the identification of 137148, 91995 and 203400 contigs for cultivars Nattu, Erode and Mysore respectively. They reported that sequence similarity search identified the presence of curcumin synthase but they could not detect significant differences in the expression level of transcripts among three cultivars at the transcriptome level.

Most of the genes and enzymes are active when the secondary metabolite level is in its peak and or even earlier and hence transcriptomes were developed at this stage to elucidate the pathway in many non- model plants (Tao et al., 2011; Wu et al., 2016). A full length cDNA library is a requisite for functional annotation of genes (Suzuki et al., 1997;

Wiemann et al., 2003). Due to the non availability of whole genome sequence of turmeric, gene transcripts and transcript variants need to be cloned to characterise gene function and regulation. The availability of complementary DNA (cDNA) copies of mRNA's provides an extremely powerful tool for expression studies and structure and organisation of genes. The availability of full length cDNA sequences helps in identifying initiation, coding and termination sequences and also can help in developing probes for mining and characterisation of corresponding genes from the chromosomal DNA. Full length cDNA will provide the most robust method to validate a gene as it will demonstrate expression as cDNA simultaneously providing sequence information to confirm the gene model. (Peyret et al., 1995; Amiruddin et al., 2012; Mochida et al., 2013; Kuroshu et al., 2010). This can further be validated by expression studies *in vivo* and comparison under various experimental conditions contrasting in metabolite content.

There is a good opportunity to fully sequence and annotate genes using the transcriptome data and computer algorithm based predictions. These "hypothetical genes" can be supported by nucleic acids and protein homologues from other species/EST matches and co expression analysis to confirm their role in the pathway. A huge repertoire of genes putatively involved in curcumin biosynthesis are neither represented in the EST collections nor sequenced by public efforts. This warrants a focussed and targeted approach to clone, sequence and undertake functional validation of these hypothetical genes to pinpoint those that are the actual players in regulating the pathway. This is a complicated process due to their existence as isoforms. In order to undertake any kind of genetic manipulation for regulating expression of key genes it is imperative to identify the correct isoforms involved in the pathway. Thus full length cloning of important genes of the curcumin biosynthetic pathway can help not only in characterisation of the corresponding genes involved in curcumin biosynthesis but also due to their intact nature, can help in developing expression vector systems capable of directing synthesis of complete proteins.

The amplification of full length cDNA involving 5' and 3' UTR is essential for functional analysis of plant genes (Wiemann et al., 2003). Normally cDNA libraries are resorted to for full length gene amplification and go a long way in improving the gene annotation predictions. cDNA libraries generated by conventional methods as reported by Gubler and Hoffman (1983) contain large number of truncated clones due to premature amplification by reverse transcriptase. Added to this, due to the size preference not favouring larger fragments, the libraries possess low number of larger fragments. Hence it

is important to develop enriched library with full length cDNAs. Several methods like RNA oligo ligation to 5' mRNA (Suzuki et al., 1997), cap affinity selection through eukaryotic initation factor (Edery et al., 1995), 5' cap biotinylation and biotin affinity selection (Carninci and Hayashizaki, 1999) methods were used in the past. However, due to the laborious nature and lengthy procedures and due to requirement of large quantitites of mRNA such methods were less preferred. SMART technology on the other hand is robust and requires only a maximum 1 µg of starting mRNA (Zhu et al., 2001).

Normalization is the process by which the concentration of highly abundant transcripts was reduced and equalizes the transcript concentration in a cDNA sample to increase the identification of a rare gene (Bogdanova et al., 2010). Normalization of cDNA was found to be essential for novel/rare gene discovery (Shagin et al., 2002). Hence, in the present study normalized full length cDNA library was constructed using duplex specific nuclease combined with SMART technique from pooled tissues of turmeric harvested at 120 DAP, when the curcumin content was reported to be maximum through biochemical analysis (Zachariah et al., 1999; Neema, 2005). Duplex specific nuclease combined with SMART technique was reported as an efficient tool for full length gene amplification in *Sesamum indicum* and *Agave sisalana* sesame and sisal (Tao et al., 2011; Zhou et al., 2012). The efficiency of normalization assessed by comparing the abundance of high copy number housekeeping gene before and after normalization (Zhulidov et al., 2004) using real time PCR was found to be 4.025 and none of the clones had insert size above 3 kb. The normalization efficiency above 4 and absence of cDNA amplification above 3 kb indicate the good quality of cDNA library (Zhulidov et al., 2004). Forty clones having insert size above 1.5 kb were sequenced in one direction but only 16 clones had sequence size > 300 bp. Of them, three clones had no similarity with non redundant database in NCBI. Zhou et al. (2012) also constructed four normalized full length cDNA libraries and out of 4500 clones were sequenced; only 3875 clones gave sequence results with 1162 unique unknown genes. The sequencing of more clones may identify the genes related to curcumin biosynthesis, but it is a very costly approach.

5.3 Identification of candidate genes involved in curcumin biosynthesis

With the advent of next generation sequencing, many transcriptomes were generated from many non- model plants due to its cost-effectiveness and *de novo* assembly feature (Bhattacharyya et al., 2013; Gupta et al., 2013; Roslan et al., 2012; Pathak et al.,

2013; Torre et al., 2016). The whole genome sequence of *Curcuma* has not been determined and transcription data associated with curcuminoids biosynthesis is not available. The comparison of transcriptomes in a contrasting background of metabolite content aided in elucidating the unique secondary metabolite pathway in many plants (Roberts and Roalson, 2017; Xu et al., 2015; Jin et al., 2017). Hence the rhizome transcriptomes of *Curcuma* species with contrasting curcumin content was constructed on Illumina HiSeq 2000 platform from a high curcumin *C. longa* variety *viz.,* Megha turmeric and *C. aromatica*, a zero curcumin genotype. The transcriptome generated 3.30 and 3.71-Gb paired-end short reads of 100 nt in length from rhizomes of *C. longa* and *C. aromatica*, respectively. After removing the adaptor sequences and low-quality sequences from the raw data, high quality reads of 2.87 Gb (86 bp×2) and 3.16 Gb (85 bp×2) generated 3,43,144 transcripts and 3,81,377 transcripts respectively for *C. longa* and *C. aromatica* respectively. N50 value was 424 bp and 410 bp and average transcript length of 366.65 bp and 358.60 bp for *C. longa* and *C. aromatica* respectively (Sheeja et al., 2015).

The average GC content was 46.24% and 46.95% for *C. longa* and *C. aromatica* respectively which is comparable with that of other monocots (Smarda et al., 2014). Only those transcripts having >150 bp were used for gene expression analysis and those having Fragments Per Kilobase of transcript per Million mapped reads fragments (FPKM) ≥1 was annotated. Thus 75,468 transcripts and 79,728 transcripts of *C. longa* and *C. aromatica* respectively had expression ≥1 FPKM. However, among this, only 42,310 transcripts (56.06%) of *C. longa* and 45,651 transcripts (57.26%) of *C. aromatica* could be annotated, which is similar to the earlier report in turmeric transcriptome by Annadurai et al. (2013) where they could also annotate only 54.6% of transcripts. The transcripts without homologous hits were reported in *de novo* transcriptome assembly of other non model plants (Iorizzo et al., 2011) and this may be due to the presence of highly specific (novel genes) or non-coding cDNAs or attributed to technical/biological biases (Logacheva et al., 2011).

Assigning of Gene Ontology (GO) terms to the annotated transcripts of *C. longa* and *C. aromatica* identified ATP binding, metabolic process and integral component of membrane as the most represented terms in the three major categories; molecular function, biological processes, and cellular components respectively. Similar to our results, the terms ATP binding of molecular function and integral component of membrane of cellular component were over-represented in the shoot and root specific transcriptomes of *Bacopa*

monnieri (Jeena et al., 2017). Apart from this, in both *Curcuma* species, most of the transcripts of biological processes category fell in to metabolic process, indicating that metabolic process are active in both rhizomes, leading to the synthesis of diverse secondary metabolites. A large portion of transcripts in metabolic process were novel, suggesting the possibility of identification of novel genes related to curcumin biosynthesis. KEGG analysis also endorses the presence of vast repertoire of transcripts related to secondary metabolite biosynthesis pathway as 1443 transcripts of *C. longa* and 1541 transcripts of *C. aromatica* and was mapped to 277 and 287 KEGG pathways, respectively with 'metabolism and biosynthesis of secondary metabolites' as the major category in both the transcriptome. Among this category, 'phenylpropanoid biosynthesis' represents the largest subgroup followed by 'stilbenoid, diarylheptanoid and gingerol biosynthesis' and 'flavonoid biosynthesis'.

Although raw reads sequenced in NGS may contain substantial noise, we had in the present study focussed on specific genes of the curcuminoid pathway by selecting target sequences of our interest. The comparison of *Curcuma* transcriptomes with contrasting curcumin content was highly valuable to shed light on the curcumin biosynthesis pathway genes. All the ten reported candidate genes of curcumin biosynthesis were identified from both the transcriptomes. More than one transcript was identified from transcriptome data for each gene, suggesting that these transcripts as a part of a same gene or members of multigene families which is in accordance with Gupta et al. (2015). The polyketide synthases are reported to be key genes involved in the biosynthesis of diarylheptanoids like curcumin (Schroder, 1997). However, differential gene expression analysis showed no significant differences in the expression level of reported PKS transcripts ie., *DCS*, *CURS1, CURS2* and *CURS3*, even when a high curcumin genotype was compared with zero curcumin genotype harvested at its optimal stage of development. This motivated us to seek out other novel PKSs from this genotype. Differential expression analysis suggested that twelve PKS transcripts are up-regulated in high curcumin genotype.

Apart from the structural genes, many transcripts showed homology to transcription factors with 39 transcription factor families in *C. longa* and 36 transcription factor families in *C. aromatica*. In both *Curcuma* species; AP2-EREBP, MYB, WRKY, NAC and bHLH transcription factor families, which were reported to regulate secondary metabolite biosynthesis (Yang et al., 2012, Liu et al., 2015) were dominant. The role of transcription factor subfamily R2R3-MYBs in regulating secondary metabolite biosynthesis were

extensively studied in many plants (Shi and Xie, 2014, Gonzalez et al., 2008). Koo et al. (2013) suggested that MYBs might have a regulatory role in curcumin biosynthesis in turmeric. The presence of 406 R2R3 MYB transcripts in *C. longa* and 413 R2R3 MYB transcripts in *C. aromatica*, suggests the possibility of MYBs as transcriptional regulators in curcumin biosynthesis imparting variation in curcumin content in different collections of *C. longa* and for disparity in curcumin content of same turmeric genotype when subjected to different agro-climatic conditions. The sequence as well as pattern of differential gene expression from transcriptome data was validated by Sanger sequencing and real time PCR analysis respectively. The assembled transcripts of *CURS3*, a reported candidate gene in curcuminoid biosynthesis from *C. longa* and *C. aromatica* showed >99% identity indicated the reliability of the transcriptome data. Further, the differential expression pattern of reported candidate genes of curcuminoid biosynthesis and novel PKSs and transcription factors analyzed by real time PCR endorses the pattern of gene expression as consistent with that of DESeq analysis data from comparative transcriptome analysis.

5.4 Co-expression analysis of candidate genes and curcuminoids

Quantitative real time PCR is an extensively used method for gene expression analysis where the fluorescent dyes or probes are used and fluorescent signal was measured during DNA amplification (Valasek and Repa, 2005). It is highly sensitive, accurate and able to quantify even rare transcripts and provide reliable and rapid results. QPCR data can be subjected to absolute or relative quantification. In absolute quantification, a standard graph is plotted with known copy number to which the unknown samples are compared. However, construction of standard for each gene is time consuming and hence relative quantification enabling quantification of relative changes in gene expression is widely used where the construction of internal standard curve for each gene is not required (Pfaffl et al., 2004).

Relative quantification employs one or more reference gene whose expression is stable under selected experimental conditions and to which the expression of target gene is normalized (Pfaffl, 2001; Yuan et al., 2006) to nullify the errors that occur due to variation in the amount of starting material. The differences in RNA integrity, reverse transcriptase efficiency and cDNA sample loading can influence the quantity of starting material (Stahlberg et al, 2004). Normalization is more relevant when the samples are from different tissues, different time course and different individuals and hence normalization of

target gene has to be performed to avoid sample-to-sample and run-to-run RT-PCR variation (Pfaffl and Hageleit, 2001). MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines were proposed by Bustin et al. (2009) to facilitate the standardization of real time PCR results.

5.4.1 Reference genes for normalization of qRT PCR analysis

Actin or *GAPDH* have been used initially to normalize the data in northern blotting and semi-quantitative RT-PCR. However these genes are recently reported to be influenced by developmental and environmental factors and hence, reference genes have to be identified for each experiment (Klie and Debener, 2011). In *Brachiaria brizantha*, *EF1α* and *ubiquitin conjugating enzyme* were found to be stable in reproductive tissues, spikelets, roots and leaves, whereas for ovary tissues, the best reference genes were *ubiquitin conjugating enzyme, eukaryotic initiation factor 4A* and *EF1α* (Silveira et al., 2009). *Protein phosphatase 2A, SAND-family protein* and *ubiquitin conjugating enzyme* were reported as stable reference genes in hybrid roses (Klie and Debener, 2011) while, *GAPDH* and *ubiquitin-1* were found to be stable in leaves and roots of sugarcane and under water deficit conditions (Andrade et al., 2017). Herein, the expression stability of six genes namely *EF1α, UBIQUITIN, 18S rRNA, GAPDH, ACTIN* and *TUBULIN* was analyzed in different tissues of turmeric under three developmental stages. Three biological replicates were pooled and analyzed with three technical replicates throughout the experiment to minimize biological and technical variability (Kitchen et al., 2010). The presence of single product in qPCR amplification was confirmed by agarose gel electrophoresis and melt curve analysis and the specificity of each product was further assessed by sequencing. The absence of amplification in NoRT (No reverse transcriptase) and NTC (non template control) were monitored to ensure the specificity of target gene amplification. PCR efficiency of atleast 90% is recommended as mandatory for selection of reference genes (Svec et al., 2015). The PCR efficiency for the reference genes ranged from 92% for *TUBULIN* to 116% for *UBIQUITIN*, suggesting that the genes are suitable for gene expression analysis. Taylor et al. (2010) suggested that correlation coefficient ($r^2 > 0.980$) is desirable for qRT PCR experiments. Except *18S rRNA* ($r^2 = 0.970$), all the remaining genes had correlation coefficient > 0.980, suggesting good linear correlation.

The stability of six reference genes used was analyzed using RefFinder tool (Xie et al., 2012), which integrates four methods namely, geNorm (Vandesompele et al., 2002),

NormFinder (Andersen et al., 2004), Bestkeeper (Pfaffl et al., 2004) and delta Ct method (Silver et al., 2006). geNorm ranks the genes based on average pairwise variation (M) of a gene with all other genes and recommends the use of more than single reference gene in a particular study. M value of <1.5 indicates the stable expression. Lower the value, higher the stability. NormFinder ranks the genes based on stability value calculated from intergroup and intragroup variation of candidate reference genes. Lesser the stability value, higher the stability of reference gene. Bestkeeper computes geometric mean, arithmetic mean, minimal and maximal value, standard deviation and coefficient of variance and executes pair-wise correlation analysis of all pairs of candidate reference genes and determines the geometric mean of the best suited genes. Lesser the standard deviation, higher the stability of gene. Delta Ct method compare relative expression of pair of genes within each sample and stability was ranked based on repeatability of gene expression differences. When six reference genes (*EF1α, UBIQUITIN, 18S rRNA, GAPDH, ACTIN* and *TUBULIN)* analyzed for their stable expression in three developmental stages and four tissues in turmeric, Reffinder tool ranked *EF1α* and *UBIQUITIN* as the best reference genes with geNorm M value of 1.025, indicating stable expression. *18S rRNA* and *GAPDH* also had M value lesser than the cut off value, which is 1.5. The delta Ct method and NormFinder identified *EF1α* as best reference gene followed by *UBIQUITIN*. However, Bestkeeper ranked *18S rRNA* as best reference gene followed by *EF1α* based on standard deviation. It has been reported that different programs resulted in identification of different stable genes to the same set of data analyzed (Klie and Debener, 2011; Mallona et al., 2010). Hence, reference genes have to be selected from a single algorithm or based on a comprehensive ranking of all the programs (Pihur et al., 2009). RefFinder tool integrating the four programs ranked *EF1α* and *UBIQUITIN* as best reference genes in turmeric based on comprehensive ranking method and were used to normalize the gene expression of target genes in the present study.

5.4.2 Expression analysis of candidate genes of curcumin biosynthesis correlated with metabolite accumulation

The gene expression analysis combined with metabolite content analysis pave the way for identification of key genes involved in secondary metabolite biosynthesis. For example, the expression of structural genes; chalcone synthase, flavanone 3-hydroxylase, dihydroflavonol 4-reductase, UFGT and transcription factors; MYB 113 and MYB 114 correlated with anthocyanin content (Park et al., 2007; Wei et al., 2011; Gonzalez et al.,

2008). Hence in order to gain insights into the molecular mechanism of curcumin biosynthesis in response to tissue specificity, developmental, genotype and environmental control at the transcription level; curcumin accumulation and expression profile of curcumin biosynthetic pathway genes in turmeric were systematically analyzed for the first time in the present study. Two turmeric genotypes used in the present study had contrasting curcumin content with high curcumin content in IISR Prathibha, a released turmeric variety of IISR and low curcumin content in Collection no. 449, an IISR turmeric germplasm collection. The curcumin content was found to be decreased when these two turmeric genotypes were planted in Coimbatore under irrigated condition, which was consistent with the earlier reports (Anandaraj et al., 2014), where a reduction in curcumin content was observed at Coimbatore in comparison with Kozhikode. The reduction in curcumin content was also observed when the turmeric genotypes are planted in grow-bags in green house with no fertilizer application. It is reported that the yield of turmeric depends on good supply of nutrients and organic matter (Srinivasan et al., 2016). Hence, the reduced curcumin content under nutrient limiting conditions in grow-bags was quite expected. Apart from this, the curcumin accumulation was found to be maximum in seedlings harvested at 120 DAP when compared with 60 and 180 DAP, which is consistent with the earlier report in IISR Prathibha by Neema (2005). The reduction of curcumin content at later stages of development might be due to increased accumulation of starch and fibre as reported by Zachariah et al. (1999).

Although it is reported that curcumin is a phenylpropanoid derivative, neither the phenylpropanoid pathway genes nor their specific isoform linked with curcumin biosysnthesis were identified in turmeric, which forms the major challenge of the present study. The transcriptomic data was highly instrumental in identifying the transcripts showing similarity to phenylpropanoid pathway genes and those having higher fold expression in *C. longa* (high curcumin content) in comparison with *C. aromatica* (zero curcumin), which were selected to design the primers. Since it was hypothesized that curcumin could be synthesized from *p*-coumaroyl CoA or/and feruloyl CoA (Ramirez-Ahmuda et al., 2006), the genes involved in biosynthesis of both CoA esters studied in other crops were analyzed in the present study. *PAL, C4H* and *4CL* are the three genes involved in the biosynthesis of *p*-coumaroyl CoA from phenylalanine and *C3H, HCT* and *COMT* are involved in the synthesis of feruloyl CoA from *p*-coumaroyl CoA (Nguyen et al., 2016). Since polyketide synthases were suggested to be involved in curcumin

biosynthesis (Schroder, 1997), four PKS identified from turmeric (*DCS, CURS1, CURS2* and *CURS3*) and other novel PKS transcripts showing higher expression in *C. longa* was opted for analysing their role in curcumin biosynthesis. The transcription factors like MYB, WRKY etc. were also reported to be involved in the regulation of secondary metabolite biosynthesis in many crops (Yang et al., 2012; Omer et al., 2013; Liu et al., 2015). The expression profile of two MYBs and a WRKY transcript were also analyzed for the first time in turmeric in the present study.

5.4.3 Spatio-temporal, genotypic and environmental role in expression of curcumin biosynthetic genes

In the present study, the transcript level of *C4H, 4CL, DCS, CURS1, CURS3* and *ClPKS a-e* correlated with curcumin content under all experimental conditions and suggest that these genes may be a part of curcumin or concurrent biosynthetic pathway. Genes associated with identical metabolic pathways are often co-expressed (Ihmels et al., 2004), so that they can catalyze linear chain of reactions. This is evident in many secondary metabolic pathways (Hamberger et al., 2007). Identification of such co-expressed genes will provide useful information to regulate production of metabolite. The expression level of *PAL, C3H* and *COMT* were higher in IISR Prathibha when compared with Collection no. 449 and in IISR Prathibha of Koz-F when compared with Koz-GH and Cbe-F. The expression of these genes were found to be slightly higher at 60 DAP than 120 DAP, however the degree of fold change was minimal, suggesting a stable expression of these genes at 60 and 120 DAP and hence may be part of same biosynthetic pathway. The expression pattern of *PAL* in leaves correlates with the earlier radiotracer studies by Neema (2005) in turmeric leaves of IISR Prathibha, where a higher activity of 2-3 fold change of PAL was observed at 60 and 90 DAP and then reduced at 120 DAP. The concentration of *p*-coumaric acid and ferulic acid was reported to be higher in leaves than rhizomes whereas caffeic acid was higher in rhizomes (Neema, 2005). In dicots, PAL deaminates phenylalanine to cinnamic acid and is further acted upon by C4H to produce *p*-coumaric, however in monocots like maize, PAL has tyrosine ammonia lyase (TAL) activity and thus tyrosine can be directly converted to *p*-coumaric acid without the involvement of C4H (Rosler et al., 1997). In the present study, turmeric being a monocot, if PAL has TAL activity, tyrosine can be directly converted to *p*-coumaric acid and could be a reason for higher concentration of *p*-coumaric acid in leaves. If so, the higher transcript level of *C4H* involved in the conversion of cinnamic acid to *p*-coumaric acid in rhizomes need to be

explored. C3H and HCT are involved in the synthesis of caffeic acid from *p*-coumaric acid and COMT converts caffeic acid to ferulic acid (Fraser and Chapple, 2011). The expression of *C3H* was found to be similar in both leaves and rhizomes, whereas, the expression of *HCT* was higher in rhizomes, which was consistent with the increased concentration of caffeic acid in rhizomes than in leaves (Neema, 2005). The transcript level of COMT was higher in leaves at 60 DAP and similar in both the tissues at 120 DAP, suggesting its putative role in accumulation of ferulic acid in leaves (Neema, 2005). In the present study, among the phenylpropanoid biosynthetic genes, *PAL, C4H, 4CL, C3H* and *COMT* showed correlation with curcumin content, suggesting their putative role in the same or concurrent pathway. No correlation was observed with expression of *HCT*, suggesting that a different isoform of HCT may be involved in curcumin biosynthesis, since *HCT* is encoded by multigene families in plants (Nguyen et al., 2016). The suppression of these phenylpropanoid biosynthetic genes was reported to reduce the secondary metabolite content in *Arabidopsis*. It is also to be noted that different isoforms of the same gene is reported in different tissues (Fraser and Chapple, 2011) and hence a detailed study of multiple isoforms of each gene is crucial.

Among the reported PKS (Katsuyama et al., 2009a, b), the expression profile of *DCS, CURS1* and *CURS3* correlated with curcumin content with highest correlation observed for *CURS1*. The transcript level of *CURS1* was significantly higher in rhizomes harvested at 120 DAP and in turmeric samples with high curcumin content under all experimental conditions. The novel transcripts *ClPKS a-d* showed significant correlation with curcumin content and its intrinsic value of relative expression was higher than *CURS1* in all the experiments of the present study. However, the expression of *CURS2*, *ClPKS e-g*, *WRKY* and *bHLH* did not correlate with curcumin content in all the experiments of the present study. *MYB4* is reported as a repressor of phenylpropanoid biosynthesis (Jin et al., 2000) and in the present study, the transcript level of *MYB4* was higher in rhizomes of IISR Prathibha harvested at 120 DAP from Cbe-F and Koz-GH, where a reduction in curcumin content was observed, suggesting its putative role as a repressor in curcumin biosynthesis. Previous study of these genes in *C. aromatica* which is a zero curcumin genotype showed that these structural genes are not missing but down-regulated, suggesting that the expression of these genes might be influenced by some regulatory mechanisms (Sheeja et al., 2015).

In the current study we have utilised a combination of RNA-seq method based on NGS technique and analysis of metabolite content and quantitative expression levels of genes to investigate the pathway. The complexity of secondary metabolite biosynthesis pathway requires not only the genetic analysis but also simultaneous determination of metabolite contents as evident from our study. According to our observations, the transcriptome analysis is not sufficient to decode the complex network of biosynthesis, especially when many of these genes exist as isoforms, it is imperative to identify the correct isoform of a particular gene that actually is responsible for the metabolite synthesis. Cloning and sequence characterisation is a plausible strategy for identifying the isoforms.

5.5. Cloning full length gene coding regions of major PKSs involved in curcumin biosynthesis

PKSs are reported to be involved in curcumin biosynthesis (Schroder, 1997; Katsuyama et al., 2009a,b). The first committed step of specific secondary metabolism from phenylpropanoid derivatives is catalyzed by PKS like chalcone synthase in flavonoid biosynthesis (Sun et al., 2015), stilbene synthase in stilbene biosynthesis (Lu et al., 2012), olivetol synthase in cannabinoid biosynthesis (Taura et al., 2009), orcinol synthase in orsellinic acid biosynthesis (Taura et al., 2016) etc. DCS, CURS1, CURS2 and CURS3 are the major polyketide synthases reported to be involved in curcumin biosynthesis in *C. longa* (Katsuyama et al., 2009a, b). In the present study, the coding region of DCS, CURS1 and CURS3 amplified from the genomic DNA of IISR Prathibha resulted in the amplification of PCR products with higher amplicon size than reported from respective PCR products amplified from cDNA (Katsuyama et al., 2009a, b), suggesting the presence of introns in these genes. The sequencing results showed that *DCS* had two introns whereas; *CURS1* and *CURS3* had one intron each which was in agreement with the earlier report of Kita et al. (2016).

5.6 Amplification of full length cDNA of specific genes

The amplification of full length cDNA involving 5' and 3' UTR is essential for functional analysis of plant genes (Wiemann et al., 2003). Due to the non availability of whole genome sequence of turmeric, gene transcripts and transcript variants need to be cloned to characterise gene function and regulation. The sequencing of clones developed from the normalized full length cDNA library of pooled tissues of turmeric (section 4.3) could lead to the identification of full length cDNA of genes related to curcumin

biosynthesis. However, cloning and sequencing of large number of clones is a costlier approach. The screening of cDNA library to identify the full length cDNA sequence of specific gene based on DNA probes (Sambrook et al., 1989) or PCR based pooled using specific primers (Munroe et al., 1995) or PCR based screening of plasmid cDNA library (Hoskins et al., 2005) or inverse PCR (Huang et al., 2003) were reported. The specificity of product, ease of use and amplification of unknown region including UTR regions in a single reaction makes the inverse PCR approach as a cost effective and suitable technique to screen the cDNA library for specific genes. Kuniyoshi et al. (2006) combined the inverse PCR and SMART (Switching Mechanism At 5'end of RNA Transcript) approach for amplification of full length cDNA library without cloning into a plasmid. The ds cDNA was synthesized only from the full length first strand cDNA, thus eliminating truncated products from the library with the usage of SMART technique. The ds cDNA could be stored for long duration and once synthesized, can be used as the template for amplification of multiple full length cDNAs. The method of Kuniyoshi et al. (2006) was adopted to isolate the full length cDNA of *CURS3* and *ClPKS11*.

5.6.1 Full length cDNA amplification of *CURS3*

The full length cDNA of *CURS3* amplified using inverse PCR combined with SMARTer (Switching Mechanism At 5' end of RNA Transcript) PCR technology amplified coding regions as well as both 5' and 3' UTRs, suggesting the utility of the protocol for further amplification of candidate genes from the full length ds cDNA library generated from *C. longa* in the present study. uORFs and internal ribosome entry sites of 5'UTR and polyadenylation sequences at 3' UTR were detected in *CURS3* and these features were reported to influence the efficiency of translation (Wilkie et al., 2003). The deduced amino acid coding sequence of CURS3 was found to be conserved in *C. longa* and in other *Curcuma* spp. However distinct sequence variation was observed in 3'UTR of *C. longa* and *C. aromatica* (zero curcumin genotype). The sequences within 3'UTR were reported to modulate translation (Wilkie et al., 2003). A detailed study is required to assume whether indels in 3'UTR affect the curcumin biosynthesis as it is conserved in *C. longa* irrespective of variations in curcumin content but absent in genotypes devoid of curcumin such as *C. aromatica* and *C. caesia*.

5.6.2 Full length cDNA amplification of novel *ClPKSa*

The inverse PCR combined with SMARTer PCR technology protocol was used to amplify the full length cDNA sequence of *ClPKSa* and the full length novel sequence was designated as *ClPKS11*. The sequence similarity showed that the four transcripts (*ClPKS a-d*) which showed higher expression in *C. longa* in comparison with *C. aromatica* in the study may be the truncated fragments of the same gene or its isoforms (Upadhyay et al., 2014). The structural and *in silico* functional characterization of *ClPKS11* is detailed in the subsequent sections.

5.7 Structural characterization of *ClPKS11*

Our primary goal was cloning, sequencing and functional validation of the full length cDNAs of this hypothetical gene *ClPKS11*. In fact this is a challenging task as there is no evidence for the validity of this gene model nor the functions of the encoded proteins or for the transcription of the gene supported by public databases. However, the homologue of this gene with 72% similarity could be identified by blast as CURS3, which to a certain extent indicates its putative involvement in curcumin biosynthesis. The main reasons is the apparent lack of expression profiles of specific tissues, very low expression of genes and their restriction to specific tissues/developmental stages that are unexplored together with invalid gene prediction and validation. Thus it is expected that this effort to clone the full length cDNA, confirmation of the gene model using computer algorithms and validation of its function *in vivo* by expression patterns and co expression analysis *vis a vis* metabolite content will go a long way in decoding the pathway and its manipulation for creating the desired phenotype.

As detailed in section 4.8, the cDNA sequence analysis of *ClPKS11* revealed the presence of an ORF of 1176 bp that corresponded to a deduced protein sequence of 391 amino acid residues with molecular weight of 42.9 kDa which lies in the range of other reported plant PKS (Roslan et al., 2013; Resmi et al., 2013; Taura et al., 2009; Taura et al., 2016; Radhakrishnan et al., 2009; Abe et al., 2004). uORF and polyadenylation sequences, the features reported to influence translation efficiency (Wilkie et al., 2003) were also detected in *ClPKS11* as in CURS3 (Section 4.8.2 of the present study). The sequencing of gDNA using the same end to end primer revealed that the coding region of *ClPKS11* is interrupted by one intron of 91 bp long. The presence of one intron in the coding region of PKS was also found in CURS1, CURS2 and CURS3 (Kita et al., 2016, Section 4.8.1 of the

present study) and in other plant PKS (Resmi et al., 2013). The deduced amino acid sequences of PKS with typical chalcone forming reaction are reported to have >85% identity (Brand et al., 2006). However, ClPKS11 showed only 80% identity to a predicted protein (CURS2 like polypeptide of *Musa accuminata*) and 72% identity to CURS3, indicating that it is different from the already reported PKS and is thus a novel protein.

Chalcone synthase catalyzes sequential condensations of *p*-coumaroyl-CoA with three molecules of malonyl-CoA to produce the tetraketide naringenin chalcone (Ferrer et al., 1999). Steric alterations of amino acid residues in active site is observed in functionally divergent 2-pyrone synthase from *Gerbera hybrida* (T197M/G256L/S338V), owing to increased steric bulk of three active site residues occluding the substrate binding pocket, thus reducing cavity volume to accept only smaller starter substrates (Jez et al., 2000). Similar replacements were observed in quinolone synthase from *Aegle marmelos* (T132S/S133A/F265V) (Resmi et al., 2013), olivetol synthase from *Cannabis sativa* (T132A/T194M/T197L) (Taura et al., 2009), aloesone synthase from *Rheum palmatum* (T197A/G256L/S338T) (Abe et al., 2006) and in three PKS from *Aloe arborescens* namely, pentaketide chromon synthase (T197M/G256L/S338V) (Abe et al., 2005a), heptaketide synthase (T197A/G256L/S338I) (Mizuuchi et al., 2009) and octaketide synthase (T197G/G256L/S338V) (Abe et al., 2005b).

The catalytic triad residues of PKS; Cys 164, His 303 and Asn 336 catalyzing chain elongation reaction (Taura et al., 2009) was conserved in ClPKS11, suggesting ClPKS11 belongs to PKS superfamily and can catalyze polyketide synthesis. The six residue loop (Thr132-Met132), *cis*-peptide bond (Met137 and Pro138) and gatekeepers (Phe215 and Phe265) were found to be conserved in ClPKS11 as in well explored PKS, chalcone synthase. GFGPG loop which is highly conserved in plant type III PKS was found to be replaced by AFGPG in ClPKS11. However, the first residue of the loop, glycine was found to be substituted with alanine (A) or serine (S) in other PKSs (Mallika et al., 2011). AFGPG loop was also found in curcuminoid synthase of *Oryza sativa*. The catalytic machinery of polyketide biosynthesis explained based on the studies of CHS from *Medicago sativa* showed that typical chalcone forming reaction depends on the conservation of amino acid residues constituting the catalytic triad, cyclization pocket, substrate binding pocket and geometry shapers determining the overall architecture of the enzyme (Ferrer et al., 1999). As reported by Resmi and Sonia (2012), the amino acid residues in the cyclization pocket, substrate binding pocket and geometry shapers of

ClPKS9 were conserved as that of CHS, suggesting it as a chalcone forming PKS. The residues in the cyclization pocket of CHS include Thr132, Met137, Ile254, Gly256 and Pro375 (Ferrer et al., 1999) whereas in ClPKS11, Ile254 (aliphatic amino acid) was replaced by Val (aliphatic amino acid). Similar substitution of Ile254 by Val was also found in other reported PKS of turmeric namely, DCS, CURS1, CURS2, CURS3 and ClPKS10 (**Table 11**) and in ZoPKS, a non chalcone forming PKS of *Zingiber officinale* (Radhakrishnan et al., 2009). However, in these non chalcone forming PKS of *Zingiberaceae* family, Thr132 was replaced by Ala in DCS of turmeric and by Ile in CURS1, CURS2, CURS3 and ClPKS10 of turmeric and in ZoPKS of ginger in addition to Ile254 substitution by Val. Mutant studies from quinolone synthase from *Aegle marmelos* and acridone synthase from *Ruta graveolens* revealed that Thr132 and Ser133 play a crucial role in substrate specificity. These two amino acids were replaced by Ser132 and Ala133 in both quinolone synthase and acridone synthase, which might have altered the steric geometry of the active site, thus enabling them to accept more bulky N-methylanthraniloyl-CoA than *p*-coumaroyl CoA in CHS as substrate (Resmi et al., 2013; Lukacin et al., 2001). In all non chalcone forming PKS of *Zingiberaceae* family, the Ser 133 is conserved as that of CHS. In ClPKS11, Thr132 was conserved as that of CHS, where other reported non chalcone forming PKS of *Zingiberaceae* family showed variation. In CURS1-3 and ZoPKS, T132 is replaced by more bulky Ile, whereas in DCS, Thr132 is substituted with less bulky Ala. As mentioned above, these amino acid alterations at 132nd position might have a role in substrate binding, thus suggesting the novelty of ClPKS11. In all non chalcone forming PKS of *Zingiber officinale* family, including ClPKS11, Ile254 was replaced by Val.

The substrate binding residues of CHS comprise Ser133, Glu192, Thr194, Thr197 and Ser338 (Ferrer et al., 1999) while in ClPKS11, hydroxylic amino acids namely, Thr194, Thr197 and Ser338 are replaced by amidic Asn, aliphatic Ile and sulphur conataining Met respectively. However, in other non chalcone forming PKS of *Zingiberaceae* family namely, DCS, CURS1, CURS2, CURS3 and ClPKS10 of turmeric and ZoPKS of ginger, hydroxylic amino acids Thr197 and Ser338 were replaced by hydroxylic amino acid Ser and amidic Glu respectively. Thus ClPKS11 possess novelty in its amino acid residues of substrate binding cavity. The substitution in the amino acid residue at 197th position is considered to influence the functionality of PKS and thus generating diversity in PKS. For example, site directed mutagenesis revealed that Thr197

determines the polyketide chain length and product specificity. The bulkiness of the amino acid at 197th position plays a crucial role in the formation of length of the polyketide as in T197G substitution, in which Thr is substituted with less bulky Gly leading to octaketide biosynthesis in *Aloe arborescens* (Abe et al., 2005b). In ClPKS11, Thr197 is replaced by more bulky Ile, while in other non chalcone forming PKS of *Zingiberaceae* family, Thr197 is substituted by less bulky serine, which may influence the length of polyketide and product specificity. Ser338 was reported to be involved in extension of polyketide as in CHS mutant S338V, in which Ser is replaced by more bulky Val, trace amounts of octaketides synthesized from eight units of malonyl CoA was observed (Abe et al., 2006). In ClPKS11, Ser338 is replaced by bulky Met, while in other non chalcone forming PKS of *Zingiberaceae* family, Ser338 is substituted by Gln. Gln is more bulkier than Ser but less bulky than Met. The amino acid alteration at 338th position in ClPKS11 and other non chalcone forming PKS of *Zingiber officinale* may also influence the length of polyketide and product specificity.

The geometry shapers maintaining the overall architecture of PKS including Pro138, Gly163, Gly167, Leu214, Asp217, Gly262, Pro304, Gly305, Gly306, Gly335, Gly374, Pro375 and Gly376 were conserved in CHS- like proteins (Ferrer et al., 1999). However, three amino acids were altered in all the non chalcone forming PKS of *Zingiberaceae* at 163rd, 214th and 306th position, in which ClPKS11 had conserved residues as CURS1, CURS2, CURS3, ClPKS10 and ZoPKS at 163rd and 306th position. At 214th position, Leu is altered with Gly in CURS1, CURS2, CURS3, ClPKS10 and ZoPKS, whereas in ClPKS11, Leu is substituted with Ala, thus suggesting the novelty of ClPKS11 in its function. Mutational studies showed that increase in the bulkiness of the amino acids at active site was reported to reduce the cavity volume of the PKS (Resmi et al., 2013; Lim et al., 2016). When compared with CURS1, in ClPKS11, bulky residues were substituted at several positions including S197I, Q338M and G214A which might be a reason for slight reduction in the cavity volume of ClPKS11 (1690 Å^3) than CURS1 (1805 Å^3).

Phylogenetic analysis showed that ClPKS11 is located together with type III PKS of *M. acuminata* in a cluster containing non-CHS enzymes including isoforms of curcumin synthase, diketide CoA synthase, curcuminoid synthase of *O. sativa* etc. These non chalcone forming PKS are reported be involved in the biosynthesis of species specific metabolites (Radhakrishnan et al., 2009) and thus ClPKS11, grouping with a non-CHS

enzymes suggest that it may have a crucial role in the biosynthesis of curcumin, the major metabolite of turmeric.

ClPKS11 had alpha helix structure as the dominant structural component and is localized in cytoplasmic matrix; both the features are in accordance with the earlier report of other PKS (Mallika et al., 2011). The quality of the modeled protein structure has to be validated using most reliable Ramachandran plot (Ramachandran et al., 1963). One humdred % of non-glycine and non-proline residues of ClPKS11 showed good fit with reference geometry suggesting that the modeled protein structure is of good quality. Molecular docking performed with feruloyl CoA, *p*-coumaroyl CoA, feruloylydiketide CoA and *p*-coumaroyldiketide CoA as substrates (ligands) for ClPKS11 and CURS1 showed that all the substrates docked to the enzyme active site of ClPKS11 and CURS1 with lower MolDock score for ClPKS11. Molegro Virtual Docker uses an energy based scoring function; lower energy scores represent better protein-ligand bindings compared with higher energy values (Thomsen and Christensen, 2006). MolDock scoring function of all the four substrates was lower for ClPKS11 when compared with CURS1, suggesting their higher affinity of these substrates for ClPKS11. This revealed that ClPKS11 may be involved in curcumin biosynthesis. Among the substrates, feruloyl CoA exhibited the lowest docking score, suggesting higher affinity of feruloyl CoA for ClPKS11 and CURS1. This is agreeing with *in vitro* studies by Katsuyama et al. (2009a); where curcumin was detected even from starter substrates (*p*-coumaroyl CoA/feruloyl CoA) and diketide CoAs when incubated with purified recombinant CURS and malonyl CoA as extender. Additionally, Arg64 of ClPKS11 was involved in either hydrogen bond or steric interaction with all the four substrates, suggesting that Arg64 could be important for substrate binding in ClPKS11. Arg(R)64 corresponds to Lys(K)62 of chalcone synthase and Lys62 was reported to be one of the conserved residue (K55, R58 and K62) in the CoA binding tunnel of chalcone synthase (Zhou et al., 2011). K55 and K62 are conserved in DCS, CURS1, CURS2 and CURS3. R58 is conserved in DCS and CURS1 while R58L and R58V substitutions were observed in CURS2 and CURS3 respectively. However, K55 and R58 residues are conserved while K62 is replaced by R64 in ClPKS11, suggesting the novelty of ClPKS11 in substrate binding properties.

5.8 Intron length polymorphism of *ClPKS11*

Introns are abundant in most eukaryotes and are ubiquitous in diverse sequences components of genes (Wu et al., 2013; Barbazuk et al., 2008; Fedorov et al., 2002). These regions are evolutionarily less conserved and show greater variability in comparison with coding sequences, which makes them advantageous as highly polymorphic genetic markers. Intron length polymorphisms (ILP) have been utilized for developing species specific (Zhao et al., 2009) or cultivar specific molecular marker (Shimada et al., 2009). Intron-spanning markers or ISM are also used for large scale genotyping applications in many crops (Badoni et al., 2016; Wang et al., 2005b; Gupta et al., 2011). These markers show high specificity, are robust and reproducible. They are co-dominant and multi-allelic with ubiquitous presence in genomes of crops. The markers could be directly associated with traits of importance.

The primers for amplification of these gene regions are designed from the exonic regions, these markers are transferrable to related species also (Gupta et al., 2012). Five ILP markers designed from four flavonoid biosynthetic genes could differentiate nine gentian cultivars and nine breeding lines (Shimada et al., 2009). ILP markers developed from flavonoid biosynthetic genes were also developed to distinguish tea (Kaundun and Matsumoto, 2003) and strawberry cultivars (Kunihisa et al., 2003). Species level discrimination in *Curcuma* spp. was reported by Kita et al. (2016) using DCS and CURSs. In the present study, subspecies level discrimination of three high curcumin and three low curcumin genotypes were analyzed based on introns of *ClPKS11*. The amino acid sequence of ClPKS11 was highly conserved in *C. longa* and in other *Curcuma* spp irrespective of their variations in curcumin content. However, variation in size as well as in sequence was observed in its intron but this could not be associated with curcumin content. This result is in agreement with earlier report of Kita et al. (2016), where they could observe sequence conservation in the coding regions of DCS, CURS1, CURS2 and CURS3 but not in case of intron. Additionally, high curcumin containing released variety of turmeric, Megha turmeric was discriminated from other released varieties; IISR Prathibha and IISR Aleppey supreme by three bp insertions. IISR Prathibha and IISR Aleppey supreme were differentiated by one base substitution (C/T). Thus in conformation with earlier studies in tea, strawberry and gentian flowers, ILP based marker could be utilised for differentiating *C. longa* genotypes and can help to protect the breeders' rights.

5.9 Pre-mRNA analysis of *ClPKS11*

No amplification was observed with pre-mRNA analysis in agarose gel electrophoresis with 25 ng of DNase treated RNA. The presence of pre-mRNA was confirmed through qRT-PCR with ~5 ng of cDNA. To confirm post-transcriptional silencing of gene expression, comparison of expression levels of pre-mRNA and mRNA is used as a preliminary tool in many studies (Morita et al., 2012). In *Petunia hybrida*, the expression of pre-mRNA and mRNA of chalcone synthase was analyzed. The pre-mRNA was amplified in both pigmented and white tissues of *P. hybrida* with same intensity indicating the transcription of the genes in both the tissues. However, mature mRNA was detected only in pigmented tissues. The absence of amplification of chalcone synthase transcripts in white tissues of *Petunia hybrida* is reported due to post-transcriptional silencing and degradation of chalcone synthase transcripts (Morita et al., 2012).

In the present study, pre-mRNA level did not correlate with variation in curcumin content. However, when expression of mRNA of candidate gene *ClPKS11* was compared between IISR Prathibha (high curcumin genotype) and Collection no. 200 (low curcumin genotype), expression level of mRNA was higher in IISR Prathibha than Collection no. 200; hence differential expression of mRNA correlated well with curcumin. This suggests the absence of post-transcriptional silencing activity and also indicates the involvement of key transcription factors in curcumin biosynthesis.

Only limited studies are reported in pre-mRNA splicing of plants. Serine/Arginine rich (SR) proteins, heterogeneous nuclear ribonucleoparticle proteins (hnRNPs) and spliceosomal proteins are reported to be involved in pre-mRNA splicing in *Arabidopsis* (Meyer et al., 2015). Serine/Arginine rich (SR) proteins are reported to be essential for constitutive and alternative splicing of pre-mRNAs (Howard and Sanford, 2015) and their expression is influenced by multiple environmental stress factors (Duque, 2011). Differentially expressing transcripts encoding these proteins were identified in *Curcuma* transcriptome (data not provided). However, a detailed study is required to understand the mechanism behind the factors regulating the splicing process of pre-mRNA of *ClPKS11*.

5.10 Identification of upstream sequence of *ClPKS11*

Transcription factors (TFs) regulate gene expression through binding to *cis*-regulatory specific sequences in the promoters of their target genes (Chang et al., 2006).

Analysis of gene expression data generated in several high throughput profiling experiments has demonstrated that genes that have a similar expression pattern are often linked and coordinate for a particular function (Vandepoele et al., 2009). For better understanding of transcriptional regulation, identification of the putative positions of the promoter and *cis*-regulatory element binding sites regulating gene expression is vital (Yamamoto et al., 2007). Integration of *cis*-regulatory elements and gene co- expression networks provide a good source for functional annotation of genes. The *cis*-acting elements involved in light, stress and hormone-responsiveness and MYB binding sites were predicted from upstream region of phenylalanine ammonia lyase of *Solenostemon scutellarioides* (Zhu et al., 2015). TATA box promoter region, transcription start site and MYB binding sites were predicted from upstream sequence of start codon of *ClPKS11*. A sequence of 495 bp upstream of start codon (translation start site) of *ClPKS11* was obtained using restricted genomic DNA and inverse PCR method and transcription start site (TSS) was observed at 201 bp. The TATA box with sequence ATATAT was observed at -33 bp position from transcription start site. The sequence of TATA box resembles that of *Brassica napus* as observed from Plant CARE database. The distance between TATA box and TSS was 33 bp which was in accordance with the report of Breathnach and Chambon (1981) and Lin et al. (2005) who suggested a distance between TATA box and TSS in most genes is about 32±7 bp. The core promoter element, CAAT box was also observed in the upstream region of ClPKS11. In addition, many *cis*-acting elements involved in light responsiveness namely; L-box, Box4, G-box, GA motif, ATCC motif, ATCT motif and TCT motif were predicted from this region, suggesting that light could play a key role in regulating the expression of ClPKS11. Besides, there were stress and hormone-responsive elements including Box-W1, DRE, HSE, TC-rich repeat, CGTCA motif, TGACG motif, GARE motif and TGA motif; indicating that the expression of *ClPKS11* may also be regulated by pathogen infection, cold and dehydration stress, heat stress, abscissic acid, gibberllin, auxin and methyl jasmaonate. Putative binding sites of R2R3-MYBs were also identified from the 5′-flanking region sequence of *ClPKS11*, including MYB4; a transcriptional repressor of phenylpropanoid biosynthesis (Shen et al., 2012). Many differentially expressing R2R3 MYBs were identified from the transcriptome (section 4.4.2). Since R2R3 MYB binding sites were predicted from the upstream sequence of *ClPKS11* and the expression of *ClPKS11* is influenced by different environmental conditions, it is hypothesized that some of the R2R3 MYBs might be regulating the expression of *ClPKS11*.

5.11 Extraction and electrophoretic separation of proteins from turmeric rhizomes

The presence of secondary metabolites interferes with extraction and separation of proteins (Rhodes, 1994). The presence of phenolics, carbohydrates, terpenoids, pigments, lipids and waxes causes streaking on the gel (Valcu and Schlink, 2006). Thus sample preparation becomes a critical step for a proteome study of turmeric rhizomes as it is being rich in carbohydrates, phenolics and terpenoids (Sasikumar, 2005; Babu et al., 1993). The phenol based protocol was reported to be superior when compared with other protocols like trichloroacetic acid/acetone, direct lysis buffer extraction in extracting proteins from banana roots (Vaganan et al., 2015), grape tissues (Marsoni et al., 2005), sugarcane stalk tissues (Amalraj et al., 2010), recalcitrant woody plants (Sebastiana et al., 2013), pigeonpea tissues (Singh et al., 2015) etc. Chokchaichamnankit et al. (2009) compared three solvent systems namely acid-acetone, phenol and trichloroacetic acid to extract protein from turmeric rhizomes and reported phenol as the better-quality solvent in extracting proteins from turmeric rhizome. In the present study, the phenol based protocol was applied to extract protein from rhizomes of high curcumin and low curcumin turmeric genotypes. The protein yield was in agreement with previous studies on protein extraction from recalcitrant tissues (Sebastiana et al., 2013; Luis et al., 2016; Rodrigues et al., 2012).

The integrity of protein is routinely assessed by SDS-PAGE coupled with Coomassie blue staining which could detect as little as 100 ng of protein (Raynal et al., 2014). The protein from rhizomes of turmeric when analyzed by SDS-PAGE generated distinct bands ranging from 20 to 66.4 kDa with no degradation, suggesting the suitability of the phenol extraction method for two dimensional gel electrophoresis (2DE). In 2DE, the proteins are separated based on their isoelectric point and molecular weight and identified using mass spectroscopy (Luis et al., 2016). In the present study, proteins were extracted from rhizomes at 120 DAP from a high curcumin genotype and a low curcumin genotype, separated on the IPG strip (pH 4-7) and SDS PAGE to generate 2DE map of these two genotypes. The visual analysis of these 2DE gels stained with Coomassie brilliant blue revealed 42 differentially expressing spots. In both the samples, high intensity spots were observed in acidic region (towards pH-4) and these results are in agreement with earlier work by Chokchaichamnankit et al. (2009) who developed the 2DE map of turmeric rhizomes from dormant and sprouting stages.

Summary and future prospects...

Summary and future prospects

Turmeric (*Curcuma longa* L.) is highly valued as a spice and known for its medicinal properties from time immemorial. Although most of the therapeutic properties of turmeric are attributed to the presence of non flavonoid polyphenol called curcumin, studies on biosynthesis of curcumin and factors affecting its biosynthesis is still in its infancy. The biosynthesis of curcumin takes place via phenylpropanoid pathway. In turmeric, curcumin is present in very low quantity and its accumulation was reported to be influenced by both genotypic as well as environmental factors, which is a major constraint for the spice industry. In order to increase the yield and stability of curcumin accumulation, there is a need to have an in-depth knowledge of both structural as well as regulatory genes of curcumin biosynthesis pathway.

In this context, the present work was undertaken to isolate and characterize the key genes/enzymes involved in curcumin biosynthesis. The turmeric rhizome being a recalcitrant tissue and in our present study extraction of functional total RNA is a prerequisite for gene expression and transcriptome analysis, four RNA isolation protocols reported from recalcitrant tissues were evaluated to extract RNA from leaf and rhizome tissues. All the protocols worked well for leaf tissues but, either completely failed or yielded low quality/quantity RNA in case of rhizome tissues. Hence a rapid SDS-acid phenol: chloroform based protocol was optimized to extract RNA from rhizomes of turmeric. The optimized protocol was compatible in extracting RNA not only from rhizomes but also from leaf, pseudostem and root tissues of turmeric. The protocol worked well in isolating total RNA from turmeric tissues stored in RNAlater and from rhizomes of related Zingiberaceae members. The extracted RNA was well-suited for gene expression analysis, ds-cDNA library construction, next generation sequencing and low molecular weight RNA isolation.

To develop a resource enriched with full length cDNA from four month old turmeric seedling, a normalized ds cDNA library was synthesized from pooled turmeric tissues. Concurrently, rhizome specific transcriptomes developed from turmeric with high curcumin content and wild turmeric (*C. aromatica*) of zero curcumin content on Illumina HiSeq 2000 platform were compared to identify the transcripts related to curcumin biosynthesis. All the reported candidate genes of curcumin biosynthetic pathway were mined from both the transcriptomes. Surprisingly, none of the reported genes of

curcuminoid biosynthetic pathway showed significant variation in expression even though the transcriptomes were developed from *Curcuma* species with contrasting curcumin content. However, the transcriptomes were rich in differentially expressed novel polyketide synthases and transcription factors. Twelve PKS transcripts up-regulated in turmeric and transcription factors belonging to thirty nine transcription factor members including 406 transcripts orthologous to R2R3 MYBs were identified from turmeric transcriptome.

Since the identification of reference gene is mandatory for accurate gene expression analysis, six reference genes were analyzed for stable gene expression in different tissues of turmeric under three developmental stages. *EF1α* and *UBIQUITIN* were selected as stable reference genes for normalizing the expression of target genes in the present study. The expression of ten reported genes of curcumin biosynthesis pathway namely; *PAL, C4H, 4CL, C3H, HCT, COMT, DCS, CURS1, CURS2* and *CURS3* and eleven differentially expressed transcripts mined from transcriptome analysis including polyketide synthases and transcription factors were analyzed in two turmeric genotypes with contrasting curcumin content from three different experimental conditions. The expression of *C4H, 4CL, DCS, CURS1, CURS3, ClPKS a-d* and *WRKY* were up-regulated in high curcumin genotype (IISR Prathibha) and correlated with curcumin content. Moreover, the expression of *C4H* and *ClPKS a-d* were significantly correlated with curcumin content. An MYB transcript orthologous to transcriptional repressor (*At* MYB4) was down-regulated in IISR Prathibha when analyzed from curcumin favourable condition (Kozhikode-field) and was up-regulated in stress conditions (Kozhikode-green house and Coimbatore-field), where the curcumin accumulation was lower, suggesting the probable role of *MYB* transcript as a transcriptional repressor.

Polyketide synthases (PKSs), being the key enzymes of curcumin biosynthesis, full length coding sequence of key PKSs in curcumin biosynthetic pathway, namely DCS, CURS1 and CURS3 were amplified from genomic DNA. Sequence analysis showed that *CURS1* and *CURS3* have an intron of 121 bp and 88 bp respectively. *DCS* had two introns of 113 bp and 85 bp. Since *ClPKSa* transcript that correlated with curcumin content and showed maximum fold change in gene expression studies is a partial fragment (326 bp), the elucidation of its full length cDNA sequence is indispensible for post-genomics studies. A protocol for amplification of full length cDNA based on inverse PCR utilizing normalized full length cDNA library synthesized from pooled tissues of turmeric was

optimized for this purpose. The efficiency of protocol was analyzed by amplifying the full length cDNA of *CURS3*. The full length coding sequence along with UTRs of *CURS3* was amplified. High degree of conservation in the coding sequence of CURS3 was observed when the sequences from three high curcumin genotypes, three low curcumin genotypes and two other *Curcuma* species with zero curcumin content; *C. aromatica* and *C. caesia* were compared. Notably, variation in the 3'UTR sequence is observed when the sequence of *CURS3* was analyzed between *C. longa* and *C. aromatica*.

The amplification of full length cDNA of *ClPKSa* based on the optimized inverse PCR protocol identified that the transcripts *ClPKS b-d* shared similarity with full length sequence, suggesting that they can be either fragments of the full length sequence named as *ClPKS11* or its isoforms. The ORF of novel *PKS*, *ClPKS11* is 1176 bp, bordered by 5'UTR of 81 bp and 3'UTR of 262 bp corresponding to a deduced protein sequence of 391 amino acids with predicted molecular mass of 42.9 kDa and pI of 6.11. The factors affecting translation efficiency namely, uORF in 5'UTR and polyadenylation sequences in 3'UTR were also identified from both *ClPKS11* and *CURS3* sequences. The amino acid sequence of ClPKS11 shared only ≤ 80% identity with reported PKSs but with conserved catalytic triad (Cys164-His303-Asn336), a feature of plant PKSs and clustered with non-CHS type III PKSs. Interestingly, ClPKS11 showed uniqueness in amino acid residues lining cyclization pocket, substrate binding cavity, CoA binding tunnel and in terms of geometry shapers; suggesting its novelty in mechanism of action. ClPKS11 was predicted to be localized in cytoplasmic matrix with alpha helix structure as dominant structural component and has a predicted cavity volume of 1690 $Å^3$.

The reported substrates of turmeric PKSs namely; feruloyl CoA, *p*-coumaroyl CoA, feruloylydiketide CoA and *p*-coumaroyldiketide CoA interacted with predicted model structure of ClPKS11 with higher binding energies. Only one amino acid residue; Arg64 which is novel in ClPKS11 interacted with all these substrates, suggesting it may have a role in substrate binding and imparting novelty of ClPKS11. As observed for CURS3, the sequences of ClPKS11 are conserved in the coding region in *Curcuma* spp but showed variations in the intron sequence, which varied from 88-101 bp. Pre-mRNA analysis of *ClPKS11* showed the absence of complete post transcriptional silencing activity as no pre-mRNA amplification was observed while mature mRNA was amplified from both turmeric genotypes of contrasting curcumin content with higher expression in high curcumin

genotype. A sequence of 495 bp upstream of coding region of *ClPKS11* was amplified from genomic DNA to identify the promoter sequence using restricted genomic DNA-inverse PCR method. TATA box and CAAT box promoter regions, transcription start site and *cis*-regulatory binding regions including those of transcription factors involved in secondary metabolism namely WRKY and MYB; especially the binding site of transcriptional repressor MYB4 were predicted from this upstream region.

A phenol based protein isolation protocol was standardized and was found to be ideal for the study of differentially expressed proteins from a low curcumin and high curcumin turmeric rhizomes using 2-dimensional gel electrophoresis. Based on the study 42 differentially expressed proteins were visually detected. The present method was ideal for the separation of the proteins in the 2^{nd} dimension as the gels were free of any streaks and had a clear background.

It is evident from the present study that biosynthesis of curcumin is a complex metabolic pathway and only a few PKSs were studied in detail till date. It is observed that Illumina sequencing can be applied as a rapid method for *de novo* transcriptome analysis of non model plants lacking genomic information. The transcriptome data generated in *Curcuma* spp. will accelerate research on the gene expression and functional genomics of turmeric. The identification and characterization of appropriate isoforms of curcumin biosynthetic pathway genes including phenylpropanoid biosynthetic genes, O-methyltransferases, reductases and hydroxylases is a prerequisite for unravelling the basic mechanism of curcumin biosynthesis. A novel polyketide synthase, ClPKS11 correlating with curcumin content was identified and its structural and *in silico* functional characterization was studied. The isoform analysis of ClPKS11 including the UTR and promoter regions in different tissues of turmeric and different genotypes with differential curcumin content and its functional characterization may pin point the role for this novel candidate in curcumin biosynthesis. This methodology may be adopted to identify, clone, characterise and determine the functional properties of other novel genes and transcripts identified by NGS.

It is evident from the study that the accumulation of curcumin is influenced by genetic, environmental and developmental factors; suggesting the existence of a regulatory mechanism involved in manipulating the expression of structural genes of curcumin biosynthetic pathway. The present study could detect the presence of MYB binding sites in

the upstream region of ClPKS11; suggesting MYBs may regulate the expression of structural genes. The transcriptomes identified several MYBs in response to stress as well as differential curcumin content. These transcription factors identified from turmeric transcriptome requires in-depth gene expression studies and coexpression analysis to identify the key players in this complicated regulatory pathway. The identification of proteins correlating with curcumin content by utilizing different model systems of contrasting curcumin content and using a gel free proteomics approach may deepen our understanding of the role of novel enzymes/transcription factors in this pathway. Thus the identification and characterization of key structural and regulatory genes in curcumin biosynthesis pathway and how the regulatory elements respond to internal and external factors and the mechanism by which they regulate structural genes needs to be investigated in future.

References...

References

Abdallah, C., Dumas-Gaudot, E., Renaut, J., & Sergeant, K. (2012). Gel-based and gel-free quantitative proteomics approaches at a glance. *International Journal of Plant Genomics*, *2012*.

Abe, I., & Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports*, *27*(6), 809–838.

Abe, I., Utsumi, Y., Oguro, S., Morita, H., Sano, Y., & Noguchi, H. (2005a). A plant type III polyketide synthase that produces pentaketide chromone. *Journal of the American Chemical Society*, *127*(5), 1362–1363.

Abe, I., Utsumi, Y., Oguro, S., & Noguchi, H. (2004). The first plant type III polyketide synthase that catalyzes formation of aromatic heptaketide. *FEBS Letters*, *562*(1–3), 171–176.

Abe, I., Oguro, S., Utsumi, Y., Sano, Y., & Noguchi, H. (2005b). Engineered biosynthesis of plant polyketides: Chain length control in an octaketide-producing plant type III polyketide synthase. *Journal of the American Chemical Society*, *127*(36), 12709–12716.

Abe, I., Watanabe, T., Lou, W., & Noguchi, H. (2006). Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. *FEBS Journal*, *273*(1), 208–218.

Abe, T., Morita, H., Noma, H., Kohno, T., Noguchi, H., & Abe, I. (2007). Structure function analysis of benzalacetone synthase from *Rheum palmatum*. *Bioorganic and Medicinal Chemistry Letters*, *17*(11), 3161–3166.

Abu-Rizq, H. A., Mansour, M. H., & Afzal, M. (2015). *Curcuma longa* attenuates carbon tetrachloride-induced oxidative stress in T-lymphocyte subpopulations. *Advanced Protocols in Oxidative Stress III*, 159-170

Accerbi, M., Schmidt, S. A., De Paoli, E., Park, S., Jeong, D. H., & Green, P. J. (2010). Methods for isolation of total RNA to recover miRNAs and other small RNAs from diverse species. *Plant MicroRNAs: Methods and Protocols*, 31-50.

Allen, R. S., Millgate, A. G., Chitty, J. a, Thisleton, J., Miller, J. a C., Fist, A. J., … Larkin, P. J. (2004). RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. *Nature Biotechnology*, *22*(12), 1559–1566.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Amalraj, R. S., Selvaraj, N., Veluswamy, G. K., Ramanujan, R. P., Muthurajan, R., Palaniyandi, M., … Viswanathan, R. (2010). Sugarcane proteomics: Establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. *Electrophoresis*, *31*(12), 1959–1974.

Amato, A., Cavallini, E., Zenoni, S., Finezzo, L., Begheldo, M., Ruperti, B., & Tornielli, G. B. (2016). A grapevine TTG2-like WRKY transcription factor is involved in regulating vacuolar transport and flavonoid biosynthesis. *Frontiers in plant science*, *7*.

American Spice Trade Association (ASTA), 1968. Official Analytical Methods, second ed. ASTA, New York.

Amiruddin, N., Lee, X. W., Blake, D. P., Suzuki, Y., Tay, Y. L., Lim, L. S., ... & Wan, K. L. (2012). Characterisation of full-length cDNA sequences provides insights into the *Eimeria tenella* transcriptome. *BMC genomics*, *13*(1), 21.

Ammon, H. P., & Wahl, M. A. (1991). Pharmacology of Curcuma longa. *Planta medica*, *57*(01), 1-7.

Anand, S., Prasad, M. V. R., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z., & Mohanty, D. (2010). SBSPKS: Structure based sequence analysis of polyketide synthases. *Nucleic Acids Research*, *38*(SUPPL. 2), 487–496.

Anandaraj, M., Prasath, D., Kandiannan, K., Zachariah, T. J., Srinivasan, V., Jha, A. K., … Maheswari, K. U. (2014). Genotype by environment interaction effects on yield and curcumin in turmeric (*Curcuma longa* L.). *Industrial Crops & Products*, *53*(January 2014), 358–364.

Andersen, C., Jensen, J., & Orntoft, T. (2004). Normalization of real time quantitative reverse transcription PCR data : A model based variance estimation approach to identify genes suited for normalization, Applied to bladder and colon cancer data sets. *Cancer Research*, *64*, 5245.

Andrade, L. M., Brito, M., Junior, R. P., Marchiori, P. E. R., Nóbile, P. M., Martins, A. P. B., ... Creste, S. (2017). Reference genes for normalization of qPCR assays in sugarcane plants under water deficit. *Plant Methods*, *13*(1), 28.

Annadurai, R. S., Neethiraj, R., Jayakumar, V., Damodaran, A. C., Rao, S. N., Katta, M. A. V. S. K., … Mugasimangalam, R. C. (2013). *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLoS ONE*, *8*(2).

Aryal, U. K., Xiong, Y., McBride, Z., Kihara, D., Xie, J., Hall, M. C., & Szymanski, D. B. (2014). A proteomic strategy for global analysis of plant protein complexes. *The Plant Cell*, *26*(10), 3867–82.

Austin, M. B., & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*, *20*(1), 79–110.

Awasthi, P., Lakshmi Jamwal, V., Kapoor, N., Rasool, S., & Bahler, J. (2016). Homology modeling and docking study of chalcone synthase gene (CfCHS) from *Coleus forskohlii*. *Cogent Biology*, *2*(1), 1175332.

Babu, K. N., Sasikumar, B., Ratnambal, M. J., George, J., & Ravindran, P. N. (1993). Genetic variability in turmeric (*Curcuma longa* L.). *The Indian Journal of Genetics and Plant Breeding*, *53*(1), 91-93.

Badoni, S., Das, S., Sayal, Y. K., Gopalakrishnan, S., Singh, A. K., Rao, A. R., ... & Tyagi, A. K. (2016). Genome-wide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice. *Scientific reports*, *6*, 23765.

Bais, H. P., Vepachedu, R., Gilroy, S., Callaway, R. M., & Vivanco, J. M. (2003). Allelopathy and exotic plant invasion: From molecules and genes to species interactions. *Science*, *301*(5638), 1377–1380.

Barbazuk, W. B., Fu, Y., & McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome research*, *18*(9), 1381-1392.

Baxter, H. L., & Stewart Jr, C. N. (2013). Effects of altered lignin biosynthesis on phenylpropanoid metabolism and plant stress. *Biofuels*, 4(6), 635-650.

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., & Legrand, M. (2007). Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *The Plant Cell*, *19*(1), 148–62.

Bhat, W. W., Dhar, N., Razdan, S., Rana, S., Mehra, R., Nargotra, A., ... & Lattoo, S. K. (2013). Molecular characterization of UGT94F2 and UGT86C4, two glycosyltransferases from *Picrorhiza kurrooa*: comparative structural insight and evaluation of substrate recognition. *PLoS One*, *8*(9).

Bhattacharyya, D., Sinha, R., Hazra, S., Datta, R., & Chattopadhyay, S. (2013). *De novo* transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, *Podophyllum hexandrum*. *BMC Genomics*, *14*(1), 748.

Bhuiyan, N. H., Selvaraj, G., Wei, Y., & King, J. (2009). Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. *Journal of Experimental Botany*, *60*(2), 509–521.

Bimboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research*, *7*(6), 1513-1523.

Birtic, S. & Kranner, I. (2006). Isolation of high-quality RNA from polyphenol-, polysaccharide- and lipid-rich seeds. *Phytochemical Analysis*, *17*(3), 144–148.

Bogdanova, E. A., Shagin, D. A., & Lukyanov, S. A. (2008). Normalization of full-length enriched cDNA. *Molecular BioSystems*, *4*(3), 205-212.

Bogdanova, E. A., Shagina, I., Barsova, E. V., Kelmanson, I., Shagin, D. A., & Lukyanov, S. A. (2010). Normalizing cDNA libraries. *Current protocols in molecular biology*, 5-12.

Bogs, J., Downey, M. O., Harvey, J. S., Ashton, A. R., Tanner, G. J., & Robinson, S. P. (2005). Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiology*, *139*(2), 652–63.

Bourgaud, F., Gravot, A., Milesi, S., & Gontier, E. (2001). Production of plant secondary metabolites : a historical perspective. *Plant science*, *161*, 839–851.

Brand, S., Hölscher, D., Schierhorn, A., Svatoš, A., Schröder, J., & Schneider, B. (2006). A type III polyketide synthase from *Wachendorfia thyrsiflora* and its role in diarylheptanoid and phenylphenalenone biosynthesis. *Planta*, *224*(2), 413–428.

Breathnach, R., & Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annual Reviews of Biochemistry*, *50*, 349–386.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., ... & Vandesompele, J. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, *55*(4), 611-622.

Carninci, P., & Hayashizaki, Y. (1999). High-efficiency full-length cDNA cloning. *Methods in enzymology*, *303*, 19-44.

Chakravorti, A. K. (1948). Multiplication of chromosome numbers in relation to speciation in Zingiberaceae. *Science and Culture*, *14*, 137-140.

Champagne, M. M., & Kuehnle, A. R. (2000). An effective method for isolating RNA from tissues of *Dendrobium*. *Lindleyana*, *15*(3), 165-168.

Chang, L. W., Nagarajan, R., Magee, J. A., Milbrandt, J., & Stormo, G. D. (2006). A systematic model to predict transcriptional regulatory mechanisms based on overrepresentation of transcription factor binding profiles. *Genome research*, *16*(3), 405-413.

Chang, S., Puryear, J., & Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*, *11*(2), 113–116.

Chempakam, B., & Parthasarathy, V. A. (2008). Turmeric. *Chemistry of Spices*, 97.

Chen, L., Huang, Y., Xu, M., Cheng, Z., Zhang, D., & Zheng, J. (2016). iTRAQ-Based Quantitative Proteomics Analysis of Black Rice Grain Development Reveals Metabolic Pathways Associated with Anthocyanin Biosynthesis. *PloS one*, *11*(7), e0159238.

Chin, D., Huebbe, P., Pallauf, K., & Rimbach, G. (2013). Neuroprotective properties of curcumin in Alzheimer's disease - Merits and limitations. *Current Medicinal Chemistry, 20*(32), 3955–3985.

Chokchaichamnankit, D., Subhasitanont, P., Paricharttanakul, N. M., Sangvanich, P., Svasti, J., & Srisomsap, C. (2009). Proteomic alteration during dormant period of *Curcuma longa* rhizomes. *Journal of Proteomics & Bioinformatics*, *2*(9), 380–387.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry*, *162*(1), 156-159.

Chu, H. Y., Wegel, E., & Osbourn, A. (2011). From hormones to secondary metabolism: The emergence of metabolic gene clusters in plants. *Plant Journal*, *66*(1), 66–79.

Coelho, C. P., Costa Netto, A. C., Colasanti, J., & Chalfun-Junior, A. (2013). A proposed model for the flowering signaling pathway of sugarcane under photoperiodic control. *Genetics and Molecular Research*, *12*(2), 1347–1359.

Cultrone, A., Cotroneo, P. S., & Recupero, G. R. (2010). Cloning and molecular characterization of R2R3-MYB and bHLH-MYC transcription factors from *Citrus sinensis*. *Tree genetics & genomes*, *6*(1), 101-112.

Dash, P. K. (2013). High quality RNA isolation from polyphenol-, polysaccharide-and protein-rich tissues of lentil (Lens culinaris). *3 Biotech*, *3*(2), 109-114.

Deepa, P., & Yusuf, A. (2016). Influence of different host associations on glutamine synthetase activity and ammonium transporter in *Santalum album* L. *Physiology and Molecular Biology of Plants*, *22*(3), 331–340.

Desgagne-Penix, I., Khan, M. F., Schriemer, D. C., Cram, D., Nowak, J., & Facchini, P. J. (2010).Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC plant biology*, *10*(1), 252.

Dibyendu, D. M. (2015). A brief review on plant type III polyketide synthases, an important group of enzyme of secondary metabolism. *Research Journal of Recent Sciences*, *4*(10), 138–147.

Dixit, D., Srivastava, N. K., & Sharma, S. (2002). Boron deficiency induced changes in translocation of $^{14}CO_2$-photosynthate into primary metabolites in relation to essential oil and curcumin accumulation in turmeric (*Curcuma longa* L.). *Photosynthetica*, *40*(1), 109-113.

Dong, H. P., Williams, E., Wang, D. Z., Xie, Z. X., Hsia, R. C., Jenck, A., ... & Place, A. (2013). Responses of *Nannochloropsis oceanica* IMET1 to long-term nitrogen starvation and recovery. *Plant physiology*, pp-113.

Dudareva, N., Pichersky, E., & Gershenzon, J. (2004). Biochemistry of plant volatiles. *Plant Physiology*, *135*(August), 1893–1902.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 116–118.

Duque, P. (2011). A role for SR proteins in plant stress responses. *Plant signaling & behavior*, *6*(1), 49-54.

Dutta, B. (2015). Study of secondary metabolite constituents and curcumin contents of six different species of genus *Curcuma*. *Journal of Medicinal Plants Studies*, *3*(5), 116–119.

Eden, T., Pieters, R., & Richards, S. (2010). Systematic review of the addition of vincristine plus steroid pulses in maintenance treatment for childhood acute lymphoblastic leukaemia-an individual patient data meta-analysis involving 5,659 children. *British Journal of Haematology, 149*(5), 722-733.

Edery, I., Chu, L. L., Sonenberg, N., & Pelletier, J. (1995). An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). *Molecular and cellular biology*, *15*(6), 3363-3371.

Ehlting, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., & Kombrink, E. (1999). Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant Journal*, *19*(1), 9–20.

Eigner, D., & Scholz, D. (1999). *Ferula asa-foetida* and *Curcuma longa* in traditional medical treatment and diet in Nepal. *Journal of ethnopharmacology*, *67*(1), 1-6.

Eloy, N., Voorend, W., Lan, W., Saleme, M. de L. S., Cesarino, I., Vanholme, R., … Boerjan, W. A. (2016). Silencing chalcone synthase impedes the incorporation of tricin in lignin and increases lignin content. *Plant Physiology*, *173*(February), 998-1016.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., … Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics,* 5.6, 1-30

Exposito, O., Bonfill, M., Onrubia, M., Jané, A., Moyano, E., Cusidó, R. M., … Piñol, M. T. (2009). Effect of taxol feeding on taxol and related taxane production in *Taxus baccata* suspension cultures. *New Biotechnology*, *25*(4), 252–259.

Fadus, M. C., Lau, C., Bikhchandani, J., & Lynch, H. T. (2016). Curcumin: An age-old anti-inflammatory and anti-neoplastic agent. *Journal of Traditional and Complementary Medicine*, 1–8.

Fakruddin, M., Chowdhury, A., Hossain, M. N., Mannan, K. S. Bin, & Mazumdar, R. M. (2012). Pyrosequencing-Principles and Applications. *International Journal of Life Science & Pharma Research*, *2*(2), L65-76.

Fan, C., Pu, N., Wang, X., Wang, Y., Fang, L., Xu, W., & Zhang, J. (2008). *Agrobacterium*-mediated genetic transformation of grapevine (*Vitis vinifera* L.) with a novel stilbene synthase gene from Chinese wild *Vitis pseudoreticulata*. *Plant Cell, Tissue and Organ Culture*, *92*(2), 197–206.

Fedorov, A., Merican, A. F., & Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proceedings of the National Academy of Sciences*, *99*(25), 16128-16133.

Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., & Noel, J. P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology*, *6*(8), 775–84.

Fraser, C. M., & Chapple, C. (2011). The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book*, 9, 1-19.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J. E., Gang, D. R., Ohlrogge, J., & Pichersky, E. (2005). Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. *The Plant Cell*, *17*(4), 1252–67.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. 1999. A sampling of the yeast proteome. *Molecular and cellular biology*. 19: 7357-7368.

Garg, V. K. (2011). Influence of sodicity on growth, yield, quality and ionic composition of turmeric (*Curcuma longa* L.). *Journal of Spices and Aromatic Crops*, *20*(1), 22–29.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., B. A. (2005). Protein identification and analysis tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607.

Ge, Q., Zhang, Y., Hua, W. P., Wu, Y. C., Jin, X. X., Song, S. H., & Wang, Z. Z. (2015). Combination of transcriptomic and metabolomic analyses reveals a JAZ repressor in the jasmonate signaling pathway of *Salvia miltiorrhiza*. *Scientific reports*, *5*(August), 14048–14061.

Ghangal, R., Raghuvanshi, S., & Sharma, P. C. (2009). Isolation of good quality RNA from a medicinal plant seabuckthorn, rich in secondary metabolites. *Plant Physiology and Biochemistry*, *47*(11), 1113-1115.

Ghawana, S., Paul, A., Kumar, H., Kumar, A., Singh, H., Bhardwaj, P. K., … Kumar, S. (2011). An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Research Notes*, *4*(1), 85.

Ghosh, S., Banerjee, S., & Sil, P. C. (2015). The beneficial role of curcumin on inflammation, diabetes and neurodegenerative disease: A recent update. *Food and Chemical Toxicology*, *83*, 111-124.

Goel, A., Kunnumakkara, A. B., & Aggarwal, B. B. (2008). Curcumin as "Curecumin": From kitchen to clinic. *Biochemical Pharmacology*, *75*(4), 787–809.

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant Journal*, *53*(5), 814–827.

Goodwin, S., Mcpherson, J. D., & Mccombie, W. R. (2016). Coming of age : ten years of next- generation sequencing technologies. *Nature Publishing Group*, *17*(6), 333–351.

Gordo, S. M. C., Pinheiro, D. G., Moreira, E. C. O., Rodrigues, S. M., Poltronieri, M. C., de Lemos, O. F., … Darnet, S. (2012). High-throughput sequencing of black pepper root transcriptome. *BMC Plant Biology*, *12*(1), 168.

Gordon, O. N., Luis, P. B., Ashley, R. E., Osheroff, N., & Schneider, C. (2015). Oxidative Transformation of Demethoxy- and Bisdemethoxycurcumin: Products, Mechanism of Formation, and Poisoning of Human Topoisomerase IIα. *Chemical Research in Toxicology*, *28*(5), 989–996.

Greco, M., Sáez, C. A., Brown, M. T., & Bitonti, M. B. (2014). A simple and effective method for high quality co-extraction of genomic DNA and total RNA from low biomass *Ectocarpus siliculosus*, the model brown alga. *PLoS ONE*, *9*(5).

Griesser, M., Pistis, V., Suzuki, T., Tejera, N., Pratt, D. A., & Schneider, C. (2011). Autoxidative and cyclooxygenase-2 catalyzed transformation of the dietary chemopreventive agent curcumin. *Journal of Biological Chemistry*, *286*(2), 1114–1124.

Gubler, U., & Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene*, *25*(2), 263-269.

Gunnaiah, R., Kushalappa, A. C., Duggavathi, R., Fox, S., & Somers, D. J. (2012). Integrated metabolo-proteomic approach to decipher the mechanisms by which wheat qtl (*Fhb1*) contributes to resistance against *Fusarium graminearum*. *PLoS ONE*, *7*(7).

Guo, N., Cheng, F., Wu, J., Liu, B., Zheng, S., Liang, J., & Wang, X. (2014). Anthocyanin biosynthetic genes in *Brassica rapa*. *BMC Genomics*, *15*(1), 426.

Guo, W. L., Chen, R. G., Gong, Z. H., Yin, Y. X., & Li, D. W. (2013). Suppression subtractive hybridization analysis of genes regulated by application of exogenous abscisic acid in pepper plant (*Capsicum annuum* L.) leaves under chilling stress. *PLoS ONE*, *8*(6).

Gupta, P., Goel, R., Agarwal, A. V., Asif, M. H., Sangwan, N. S., Sangwan, R. S., & Trivedi, P. K. (2015). Comparative transcriptome analysis of different chemotypes elucidates withanolide biosynthesis pathway from medicinal plant *Withania somnifera*. *Scientific Reports*, *5*(November), 18611.

Gupta, P., Goel, R., Pathak, S., Srivastava, A., Singh, S. P., Sangwan, R. S., … Trivedi, P. K. (2013). De novo assembly, functional annotation and comparative analysis of *Withania somnifera* leaf and root transcriptomes to identify putative genes involved in the withanolides biosynthesis. *PLoS ONE*, *8*(5).

Gupta, S. K., Bansal, R., & Gopalakrishna, T. (2012). Development of intron length polymorphism markers in cowpea [*Vigna unguiculata* (L.) Walp.] and their transferability to other *Vigna* species. *Molecular Breeding*, *30*(3), 1363–1370.

Gupta, S., Kumari, K., Das, J., Lata, C., Puranik, S., & Prasad, M. (2011). Development and utilization of novel intron length polymorphic markers in foxtail millet (*Setaria italica* (L.) P. Beauv.). *Genome*, *54*(7), 586-602.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series, 41*(41), 95-98.

Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B., & Douglas, C. J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa, Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. This article is one of a selection of papers published in the special issue on *Poplar* research in *Canadian Journal of Botany*, *85*(12), 1182–1201.

Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, *68*(22–24), 2831–2846.

Hassaninasab, A., Hashimoto, Y., Tomita-Yokotani, K., & Kobayashi, M. (2011). Discovery of the curcumin metabolic pathway involving a unique enzyme in an intestinal microorganism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(16), 6615–20.

Hassanpour, S., Maheri-Sis, N., Eshratkhah, B., & Mehmandar, F.B. (2011). Plants and secondary metabolites (Tannins): A Review. *International Journal of Forest, Soil and Erosion*, *1*(11), 47–53.

Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., … Saito, K. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry*, *280*(27), 25590–25595.

Hofmann, K. A., & Stoffel,W. (1993). TMbase-A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, *374*, 166.

Holton, T., & Cornish, E. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *The Plant Cell*, *7*(7), 1071–1083.

Hoskins, R. A., Stapleton, M., George, R. A., Yu, C., Wan, K. H., Carlson, J. W., & Celniker, S. E. (2005). Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP). *Nucleic Acids Research*, *33*(21), e185-e185.

Hossain, M. A., & Ishimine, Y. (2005). Growth, yield and quality of turmeric (*Curcuma longa* L.) cultivated on dark-red soil, gray soil and red soil in Okinawa, Japan. *Plant Production Science*, *8* (September 2004), 482–486.

Hou, P., Xie, Z., Zhang, L., Song, Z., Mi, J., He, Y., & Li, Y. (2011). Comparison of three different methods for total RNA extraction from *Fritillaria unibracteata* : A rare Chinese medicinal plant. *Journal Of Medicinal Plants*, *5*(13), 2834–2838.

Howard, J. M., & Sanford, J. R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdisciplinary Reviews: RNA*, *6*(1), 93-110.

Huang, C.C., Couch, G.S., Pettersen, E.F., & Ferrin, T.E. (1996). Chimera: An extensible molecular modeling application constructed using standard components. *Pacific Symposium on Biocomputing* 1:724.

Huang, S. H., Chen, S. H., & Jong, A. Y. (2003). Use of inverse PCR to clone cDNA ends. *Generation of cDNA Libraries: Methods and Protocols*, 51-58.

Huang, W., Khaldun, A. B. M., Chen, J., Zhang, C., Lv, H., Yuan, L., & Wang, Y. (2016). A R2R3-MYB Transcription Factor Regulates the Flavonol Biosynthetic Pathway in a Traditional Chinese Medicinal Plant, *Epimedium sagittatum*. *Frontiers in plant science*, *7*.

Ihmels, J., Levy, R., & Barkai, N. (2004). Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nature biotechnology*, *22*(1), 86.

Illuri, R., Bethapudi, B., Anandakumar, S., Murugan, S., A Joseph, J., Mundkinajeddu, D., & Agarwal, A. (2015). Anti-Inflammatory activity of polysaccharide fraction of *Curcuma longa* extract (NR-INF-02). *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Inflammatory and Anti-Allergy Agents)*, *14*(1), 53-62.

Iorizzo, M., Senalik, D. A., Grzebelus, D., Bowman, M., Cavagnaro, P. F., Matvienko, M., ... & Simon, P. W. (2011). *De novo* assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC genomics*, *12*(1), 389.

Ireson, C. R., Jones, D. J., Orr, S., Coughtrie, M. W., Boocock, D. J., Williams, M. L., ... & Gescher, A. J. (2002). Metabolism of the cancer chemopreventive agent curcumin in human and rat intestine. *Cancer Epidemiology and Prevention Biomarkers*, *11*(1), 105-111.

Jaakola, L., Määttä, K., Pirttilä, A. M., Törrönen, R., Kärenlampi, S., & Hohtola, A. (2002). Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiology*, *130*(October), 729–739.

Jayaprakasha, G. K., Rao, L. J. M., & Sakariah, K. K. (2002). Improved HPLC method for the determination of curcumin, demethoxycurcumin, and bisdemethoxycurcumin. *Journal of Agricultural and Food Chemistry*, *50*(13), 3668–3672.

Jeena, G. S., Fatima, S., Tripathi, P., Upadhyay, S., & Shukla, R. K. (2017). Comparative transcriptome analysis of shoot and root tissue of *Bacopa monnieri* identifies potential genes related to triterpenoid saponin biosynthesis. *BMC genomics*, *18*(1), 490.

Jez, J. M., Austin, M. B., Ferrer, J. L., Bowman, M. E., Schröder, J., & Noel, J. P. (2000). Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry and Biology*, *7*(12), 919–930.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., … Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal*, *19*(22), 6150–6161.

Jin, J., Sun, Y., Qu, J., syah, R., Lim, C.-H., Alfiko, Y., … Ye, J. (2017). Transcriptome and functional analysis reveals hybrid vigor for oil biosynthesis in oil palm. *Scientific Reports*, *7*(1), 439.

Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., … Møller, B. L. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology*, *8*(3 SPEC. ISS.), 280–291.

Kalra, S., Puniya, B. L., Kulshreshtha, D., Kumar, S., Kaur, J., Ramachandran, S., & Singh, K. (2013). *De novo* transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivilianum*. *PLoS One*, *8*(12), e83336.

Kambiranda, D., Katam, R., Basha, S. M., & Siebert, S. (2014). ITRAQ-based quantitative proteomics of developing and ripening muscadine grape berry. *Journal of Proteome Research*, *13*(2), 555–569.

Kansal, R., Kuhar, K., Verma, I., Gupta, R. N., Gupta, V. K., & Koundal, K. R. (2008). Improved and convenient method of RNA isolation from polyphenols and polysaccharide rich plant tissues. *Indian Journal of Experimental Biology*, *46*(12), 842–845.

Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., … Yazaki, K. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *Plant Journal*, *77*(4), 627–638.

Katsuyama, Y., Kita, T., Funa, N., & Horinouchi, S. (2009a). Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. *Journal of Biological Chemistry*, *284*(17), 11160–11170.

Katsuyama, Y., Kita, T., & Horinouchi, S. (2009b). Identification and characterization of multiple curcumin synthases from the herb *Curcuma longa*. *FEBS Letters*, *583*(17), 2799–2803.

Katsuyama, Y., Matsuzawa, M., Funa, N., & Horinouchi, S. (2007). *In vitro* synthesis of curcuminoids by type III polyketide synthase from *Oryza sativa*. *Journal of Biological Chemistry*, *282*(52), 37702–37709.

Katsuyama, Y., Matsuzawa, M., Funa, N., & Horinouchi, S. (2008). Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway. *Microbiology*, *154*(9), 2620–2628.

Kaundun, S. S., & Matsumoto, S. (2003). Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, *Camellia sinensis* (L.) O. Kuntze, and differentiation between *assamica* and *sinensis* varieties. *TAG Theoretical and Applied Genetics, 106*(3), 375–83.

Khraiwesh, B., Qudeimat, E., Thimma, M., Chaiboonchoe, A., Jijakli, K., Alzahmi, A., … Salehi-Ashtiani, K. (2015). Genome-wide expression analysis offers new insights into the origin and evolution of *Physcomitrella patens* stress response. *Scientific Reports*, *5*(October), 17434.

Kim, S., Han, L., Yu, B., Hhnke, V. D., Bolton, E. E., & Bryant, S. H. (2015). PubChem structure-activity relationship (SAR) clusters. *Journal of Cheminformatics*, *7*(1), 1–22.

Kita, T., Imai, S., Sawada, H., Kumagai, H., & Seto, H. (2008). The biosynthetic pathway of curcuminoid in turmeric (*Curcuma longa*) as revealed by ^{13}C-labeled precursors. *Bioscience, Biotechnology and Biochemistry*, *72*(7), 1789–1798.

Kita, T., Komatsu, K., Zhu, S., Iida, O., Sugimura, K., Kawahara, N., ... & Cai, S. Q. (2016). Development of intron length polymorphism markers in genes encoding diketide-CoA synthase and curcumin synthase for discriminating *Curcuma* species. *Food chemistry*, *194*, 1329-1336.

Kitchen, R. R., Kubista, M., & Tichopad, A. (2010). Statistical aspects of quantitative real-time PCR experiment design. *Methods*, *50*(4), 231–236.

Klie, M., & Debener, T. (2011). Identification of superior reference genes for data normalisation of expression studies via quantitative PCR in hybrid roses (*Rosa hybrida*). *BMC Research Notes 2011 4:1*, *4*(518), 16–90.

Koo, H. J., McDowell, E. T., Ma, X., Greer, K. A., Kapteyn, J., Xie, Z., … Gang, D. R. (2013). Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. *BMC Plant Biology*, *13*(1), 27.

Krisa, S., Téguo, P. W., Decendit, A., Deffieux, G., Vercauteren, J., & Mérillon, J. M. (1999). Production of 13C-labelled anthocyanins by *Vitis vinifera* cell suspension cultures. *Phytochemistry*, *51*(5), 651-656.

Kumar, K. R., Rao, S. N., & Kumar, N.R.S. (2015). Evaluation of turmeric (*Curcuma longa* L.) cultivars at agency areas of north coastal Andhra Pradesh. *Progressive Research-An International Journal*. 10. 2417-2420.

Kumar, V., Kumar, C. S., Hari, G., Venugopal, N. K., Vijendra, P. D., & B, G. (2013). Homology modeling and docking studies on oxidosqualene cyclases associated with primary and secondary metabolism of *Centella asiatica*. *SpringerPlus*, *2*(1), 189.

Kunihisa, M., Fukino, N., & Matsumoto, S. (2003). Development of cleavage amplified polymorphic sequence (CAPS) markers for identification of strawberry cultivars. *Euphytica*, *134*(2), 209–215.

Kuniyoshi, H., Fukui, Y., & Sakai, Y. (2006). Cloning of full-length cDNA of teleost corticotropin-releasing hormone precursor by improved inverse PCR. *Bioscience, biotechnology, and biochemistry*, *70*(8), 1983-1986.

Kuroshu, R. M., Watanabe, J., Sugano, S., Morishita, S., Suzuki, Y., & Kasahara, M. (2010). Cost-effective sequencing of full-length cDNA clones powered by a *de novo*-reference hybrid assembly. *PLoS One*, *5*(5), e10517.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685.

Laskowski, R. a., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(November), 283–291.

Lee, J., Jung, Y., Shin, J. H., Kim, H. K., Moon, B. C., Ryu, D. H., & Hwang, G. S. (2014). Secondary metabolite profiling of curcuma species grown at different locations using GC/TOF and UPLC/Q-TOF MS. *Molecules*, *19*(7), 9535–9551.

Lee, K. H., Morris-Natschke, S. L., Zhao, Y., & Musgrove, K. (2016). Chinese Herbal Medicine-Derived Products for Prevention or Treatment of Diseases Affecting Quality of Life. In *Medicinal Plants-Recent Advances in Research and Development* (pp. 1-35). Springer Singapore.

Lee, M.-H., Jeong, J.-H., Seo, J.-W., Shin, C.-G., Kim, Y.-S., In, J.-G., … Choi, Y.-E. (2004). Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. *Plant and Cell Physiology*, *45*(8), 976–984.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., … Rombauts, S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research*, *30*(1), 325–327.

Li, J., Zhao, W., Li, J., Xia, W., Lei, L., & Zhao, S. (2014). Research methods to study plant secondary metabolic pathways and their applications in the analysis of the biosynthetic pathway of stilbenes from *Polygonum multiflorum* -A review. *Plant OMICS*, *7*(3), 158–165.

Li, J.-H., Tang, C.-H., Song, C.-Y., Chen, M.-J., Feng, Z.-Y., & Pan, Y.-J. (2006). A simple, rapid and effective method for total RNA extraction from *Lentinula edodes*. *Biotechnology Letters*, *28*(15), 1193–7.

Li, W., Liu, X., & Lu, Y. (2016). Transcriptome comparison reveals key candidate genes in response to vernalization of Oriental lily. *BMC Genomics*, *17*(1), 664.

Lillo, C., Lea, U. S., & Ruoff, P. (2008). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant, Cell and Environment*, *31*(5), 587–601.

Lim, Y. P., Go, M. K., & Yew, W. S. (2016). Exploiting the biosynthetic potential of type III polyketide synthases. *Molecules*, *21*(6), 1–37.

Lin, J., Zhou, X., Pang, Y., Gao, H., Fei, J., Shen, G. A., … Tang, K. (2005). Cloning and characterization of an agglutinin gene from *Arisaema lobatum*. *Bioscience Reports*, *25*(5–6), 345–362.

Lin, X., Xiao, M., Luo, Y., Wang, J., & Wang, H. (2013). The effect of RNAi-induced silencing of FaDFR on anthocyanin metabolism in strawberry (*Fragaria×ananassa*) fruit. *Scientia Horticulturae*, *160*, 123–128.

Liu, J., Osbourn, A., & Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Molecular plant*, *8*(5), 689-708.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., … Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*, 1-11.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods*, *25*(4), 402–408.

Logacheva, M. D., Kasianov, A. S., Vinogradov, D. V., Samigullin, T. H., Gelfand, M. S., Makeev, V. J., & Penin, A. A. (2011). *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC genomics*, *12*(1), 30.

Loomis, W. D. (1974). Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Methods in Enzymology*, *31*(C), 528–544.

Lu, C., Meyers, B. C., & Green, P. J. (2007). Construction of small RNA cDNA libraries for deep sequencing. *Methods*, *43*(2), 110–117.

Lu, D., Zhao, W., & Zhao, S. (2012). Relevant enzymes, genes and regulation mechanisms in biosynthesis pathway of stilbenes. *Open Journal of Medicinal Chemistry*, *2*(2), 15–23.

Luis, I. M., Alexandre, B. M., Oliveira, M. M., & Abreu, I. A. (2016). Selection of an appropriate protein extraction method to study the phosphoproteome of maize photosynthetic tissue. *PloS One*, *11*(10), e0164387.

Lukacin, R., Schreiner, S., & Matern, U. (2001). Transformation of acridone synthase to chalcone synthase. *FEBS Letters*, *508*(3), 413–417.

Luo, M., Dang, P., Bausher, M. G., Holbrook, C. C., Lee, R. D., Lynch, R. E., & Guo, B. Z. (2005). Identification of transcripts involved in resistance responses to leaf spot disease caused by *Cercosporidium personatum* in peanut (*Arachis hypogaea*). *Phytopathology*, *95*(4), 381-387.

Lussier, F. X., Colatriano, D., Wiltshire, Z., Page, J. E., & Martin, V. J. J. (2012). Engineering microbes for plant polyketide biosynthesis. *Computational and Structural Biotechnology Journal*, *3*(4), e201210020.

Makkar, H. P. S., Norvsambuu, T., Lkhagvatseren, S., & Becker, K. (2009). Plant secondary metabolites in some medicinal plants of Mongolia used for enhancing animal health and production. *Tropicultura*, *27*(3), 159-167.

Mallika, V., Sivakumar, K. C., & Soniya, E. V. (2011). Evolutionary implications and physicochemical analyses of selected proteins of type III polyketide synthase family. *Evolutionary Bioinformatics*, *2011*(7), 41–53.

Mallona, I., Lischewsky, S., Weiss, J., Hause, B., Egea-Cortines, M., Lischewski, S., … Egea-Cortines, M. (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biology*, *10*(1), 4.

Marsoni, M., Vannini, C., Campa, M., Cucchi, U., Espen, L., & Bracale, M. (2005). Protein extraction from grape tissues by two-dimensional electrophoresis. *Vitis - Journal of Grapevine Research*, *44*(4), 181–186.

Martinez-Esteso, M. J., Martínez-Márquez, A., Sellés-Marchart, S., Morante-Carriel, J. a, & Bru-Martínez, R. (2015). The role of proteomics in progressing insights into plant secondary metabolism. *Frontiers in Plant Science*, *6*(July), 504.

Martinez-Esteso, M. J., Sellés-Marchart, S., Lijavetzky, D., Pedreño, M. A., & Bru-Martínez, R. (2011). A DIGE-based quantitative proteomic analysis of grape berry flesh development and ripening reveals key events in sugar and organic acid metabolism. *Journal of experimental botany*, *62*(8), 2521-2569.

Masek, T., Vopalensky, V., Suchomelova, P., & Pospisek, M. (2005). Denaturing RNA electrophoresis in TAE agarose gels. *Analytical Biochemistry*, *336*(1), 46–50.

Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., ... & Saito, K. (2010). AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant physiology*, *152*(2), 566-578.

Matsuda, S., Vert, J., Saigo, H., Ueda, N., Toh, H., & Akutsu, T. (2005). A novel representation of protein sequences for prediction of subcellular location using support vector machines, *Protein science, 14*, 2804–2813.

Mazumdar, A. B., & Chattopadhyay, S. (2015). Sequencing, *de novo* assembly, functional annotation and analysis of *Phyllanthus amarus* leaf transcriptome using the Illumina platform. *Frontiers in Plant Science*, *6*(1199), 1–22.

McFadden, R. M. T., Larmonier, C. B., Shehab, K. W., Midura-Kiela, M., Ramalingam, R., Harrison, C. A., ... & Ghishan, F. K. (2015). The role of curcumin in modulating colonic microbiota during colitis and colon cancer prevention. *Inflammatory bowel diseases*, *21*(11), 2483.

Mendelsohn, S. L., & Young, D. A. (1978). Inhibition of ribonuclease Efficacy of sodium dodecyl sulfate, diethyl pyrocarbonate, proteinase K and heparin using a sensitive ribonuclease assay. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, *519*(2), 461-473.

Meyer, K., Koester, T., & Staiger, D. (2015). Pre-mRNA splicing in plants: in vivo functions of RNA-binding proteins implicated in the splicing process. *Biomolecules*, *5*(3), 1717-1740.

Milobedzka, J., & Lampe, V. (1910). Zur Kenntnis des curcumins. *European Journal of Inorganic Chemistry*, *43*(2), 2163-2170.

Mizuuchi, Y., Shi, S. P., Wanibuchi, K., Kojima, A., Morita, H., Noguchi, H., & Abe, I. (2009). Novel type III polyketide synthases from *Aloe arborescens*. *FEBS Journal*, *276*(8), 2391–2401.

Mochida, K., Uehara-Yamaguchi, Y., Takahashi, F., Yoshida, T., Sakurai, T., & Shinozaki, K. (2013). Large-scale collection and analysis of full-length cDNAs from *Brachypodium distachyon* and integration with Pooideae sequence resources. *PLoS One*, *8*(10), e75265.

Monika, G., Punam, G., Sarbjot, S., & Gupta, G. D. (2010). An overview on molecular docking. *International Journal of Drug Development & Research*, 2(1):219-231

Morita, H., Kondo, S., Oguro, S., Noguchi, H., Sugio, S., Abe, I., & Kohno, T. (2007). Structural insight into chain-length control and product specificity of pentaketide chromone synthase from *Aloe arborescens*. *Chemistry & biology*, *14*(4), 359-369.

Morita, Y., Saito, R., Ban, Y., Tanikawa, N., Kuchitsu, K., Ando, T., … Nakayama, M. (2012). Tandemly arranged *chalcone synthase A* genes contribute to the spatially regulated expression of siRNA and the natural bicolor floral phenotype in *Petunia hybrida*. *Plant Journal*, *70*(5), 739–749.

Mostajeran, A., Gholaminejad, A., & Asghari, G. (2014). Salinity alters curcumin, essential oil and chlorophyll of turmeric (*Curcuma longa* L.). *Research in pharmaceutical sciences*, *9*(1), 49.

Munroe, D. J., Loebbert, R., Bric, E., Whitton, T., Prawitt, D., Vu, D., ... & Housman, D. E. (1995). Systematic screening of an arrayed cDNA library by PCR. *Proceedings of the National Academy of Sciences*, *92*(6), 2209-2213.

Nair, R. R., & Sasikumar, B. (2009). Chromosome number variation among germplasm collections and seedling progenies in Turmeric, *Curcuma longa* L. *Cytologia*, *74*(2), 153-157.

Neema, A. (2005). Investigations on the biosynthesis of curcumin in turmeric (*Curcuma longa* L.). Ph.D. Thesis, Calicut University

Nelson, K. M., Dahlin, J. L., Bisson, J., Graham, J., Pauli, G. F., & Walters, M. A. (2017). The essential medicinal chemistry of curcumin. *Journal of Medicinal Chemistry*, *60*(5), 1620–1637.

Nguyen, Q. T., Merlo, M. E., Medema, M. H., Jankevics, A., Breitling, R., & Takano, E. (2012). Metabolomics methods for the synthetic biology of secondary metabolism. *FEBS Letters*, *586*(15), 2177–2183.

Nguyen, T.-N., Son, S., Jordan, M. C., Levin, D. B., & Ayele, B. T. (2016). Lignin biosynthesis in wheat (*Triticum aestivum* L.): its response to waterlogging and association with hormonal levels. *BMC Plant Biology*, *16*(1), 28.

Njuguna, N. M., Ongarora, D. S. B., & Chibale, K. (2012). Artemisinin derivatives: a patent review (2006 - present). *Expert Opinion on Therapeutic Patents*, *22*(10), 1179–1203.

Ober, D. (2005). Seeing double: Gene duplication and diversification in plant secondary metabolism. *Trends in Plant Science*, *10*(9), 444–449.

Ogungbe, I. V., & Setzer, W. N. (2016). The Potential of secondary metabolites from plants as drugs or leads against protozoan neglected diseases-Part III: In-silico molecular docking investigations. *Molecules (Basel, Switzerland)*, *21*(10).

Olsen, K. M., Lea, U. S., Slimestad, R., Verheul, M., & Lillo, C. (2008). Differential expression of four *Arabidopsis PAL* genes; *PAL1* and *PAL2* have functional specialization in abiotic environmental-triggered flavonoid synthesis. *Journal of Plant Physiology*, *165*(14), 1491–1499.

Omer, S., Kumar, S., & Khan, B. M. (2013). Over-expression of a subgroup 4 R2R3 type MYB transcription factor gene from *Leucaena leucocephala* reduces lignin content in transgenic tobacco. *Plant cell reports*, *32*(1), 161-171.

Pan, M. H., Huang, T. M., & Lin, J. K. (1999). Biotransformation of curcumin through reduction and glucuronidation in mice. *Drug metabolism and disposition*, *27*(4), 486-494.

Park, E. S., Moon, W. S., Song, M. J., Kim, M. N., Chung, K. H., & Yoon, J. S. (2001). Antimicrobial activity of phenol and benzoic acid derivatives. *International Biodeterioration & Biodegradation*, *47*(4), 209–214.

Park, J. S., Choung, M. G., Kim, J. B., Hahn, B. S., Kim, J. B., Bae, S. C., … Cho, K. J. (2007). Genes up-regulated during red coloration in UV-B irradiated lettuce leaves. *Plant Cell Reports*, *26*(4), 507–516.

Park, J. S., Kim, J. B., Hahn, B. S., Kim, K. H., Ha, S. H., Kim, J. B., & Kim, Y. H. (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization. *Plant Science*, *166*(4), 953–961.

Pathak, S., Lakhwani, D., Gupta, P., Mishra, B. K., Shukla, S., Asif, M. H., & Trivedi, P. K. (2013). Comparative transcriptome analysis using high papaverine mutant of *Papaver somniferum* reveals pathway and uncharacterized steps of papaverine biosynthesis. *PloS one*, *8*(5), e65622.

Pelot, K. A., Mitchell, R., Kwon, M., Hagelthorn, D. M., Wardman, J. F., Chiang, A., … Zerbe, P. (2017). Biosynthesis of the psychotropic plant diterpene salvinorin A: Discovery and characterization of the *Salvia divinorum* clerodienyl diphosphate synthase. *Plant Journal*, *89*(5), 885–897.

Peng, J., Xia, Z., Chen, L., Shi, M., Pu, J., Guo, J., & Fan, Z. (2014). Rapid and efficient isolation of high-quality small RNAs from recalcitrant plant species rich in polyphenols and polysaccharides. *PLoS ONE*, *9*(5), 12–14.

Perry, R. P., & Kelley, D. E. (1972). The production of ribosomal RNA from high molecular weight precursors. *Journal of Molecular Biology*, *70*(2), 265–279.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–6.

Peyret, P., Perez, P., & Alric, M. (1995). Structure, genomic organization, and expression of the Arabidopsis thaliana aconitase gene plant aconitase show significant homology with mammalian iron-responsive element-binding protein. *Journal of Biological Chemistry*, *270*(14), 8131-8137.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), e45.

Pfaffl, M. W., & Hageleit, M. (2001). Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters*, *23*(4), 275–282.

Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnology Letters*, *26*(6), 509–15.

Pihur, V., Datta, S., & Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC bioinformatics*, *10*(1), 62.

Prasad, S., Gupta, S. C., Tyagi, A. K., & Aggarwal, B. B. (2014). Curcumin, a component of golden spice: From bedside to bench and back. *Biotechnology Advances*, *32*(6), 1053–1064.

Priyadarsini, K. I. (2014). The chemistry of curcumin: from extraction to therapeutic agent. *Molecules*, *19*(12), 20091-20112.

Purseglove, J.W. (1968). *Tropical Crops: Monocotyledons*. Longman, London.

Radhakrishnan, E. K., Sivakumar, K. C., & Soniya, E. V. (2009). Molecular characterization of novel form of type III polyketide synthase from *Zingiber officinale* Rosc. and its analysis using bioinformatics method. *Journal of Proteomics & Bioinformatics*, *2*(7), 310–315.

Raghavan, T. S., & Venkatasubban, K. R. (1943). Cytological studies in the family Zingiberaceae with special reference to chromosome number and cyto-taxonomy. *Proceedings: Plant Sciences*, *17*(4), 118-132.

Raharjo, T. J., Widjaja, I., Roytrakul, S., & Verpoorte, R. (2004). Comparative proteomics of *Cannabis sativa* plant tissues. *Journal of Biomolecular Techniques, 15*(2), 97–106.

Rahimi, S., Hasanloo, T., Najafi, F., & Khavari-Nejad, R. A. (2011). Enhancement of silymarin accumulation using precursor feeding in *Silybum marianum* hairy root cultures. *Plant OMICS*, *4*(1), 34–39.

Raja, R., & Sreenivasulu, M. (2015). Medicinal plants secondary metabolites used in pharmaceutical importance-An overview. *World Journal of Pharmacy and Pharmaceutical Sciences*, *4*(4), 436-447.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*(1), 95–99.

Ramirez-Ahumada, M. del C., Timmermann, B. N., & Gang, D. R. (2006). Biosynthesis of curcuminoids and gingerols in turmeric (*Curcuma longa*) and ginger (*Zingiber officinale*): Identification of curcuminoid synthase and hydroxycinnamoyl-CoA thioesterases. *Phytochemistry*, *67*(18), 2017–2029.

Ravindran, P. N., Babu, K. N., & Sivaraman, K. (Eds.). (2007). *Turmeric: the genus Curcuma*. CRC Press.

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories*, *13*(1), 180.

Reimold, U., Kröger, M., Kreuzaler, F., & Hahlbrock, K. (1983). Coding and 3'non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme. *The EMBO journal*, *2*(10), 1801.

Resmi, M. S., & Soniya, E. V. (2012). Molecular cloning and differential expressions of two cDNA encoding type III polyketide synthase in different tissues of *Curcuma longa* L. *Gene*, *491*(2), 278–283.

Resmi, M. S., Verma, P., Gokhale, R. S., & Soniya, E. V. (2013). Identification and characterization of a type III polyketide synthase involved in quinolone alkaloid biosynthesis from *Aegle marmelos* Correa. *Journal of Biological Chemistry*, *288*(10), 7271–7281.

Rhodes, M. J. C. (1994). Physiological roles for secondary metabolites in plants: some progress, many outstanding problems. *Plant Molecular Biology*, *24*(1), 1–20.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(1), 276–277.

Roberts, W. R., & Roalson, E. H. (2017). Comparative transcriptome analyses of flower development in four species of *Achimenes* (Gesneriaceae). *BMC Genomics*, *18*(1), 240.

Rodrigues, E. P., Torres, A. R., da Silva Batista, J. S., Huergo, L., & Hungria, M. (2012). A simple, economical and reproducible protein extraction protocol for proteomics studies of soybean roots. *Genetics and Molecular Biology*, *35*(SUPPL.1), 348–352.

Rohde, A., Morreel, K., Ralph, J., Goeminne, G., Hostyn, V., De Rycke, R., … Boerjan, W. (2004). Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *The Plant Cell*, *16*(10), 2749–2771.

Rombel, I. T., Sykes, K. F., Rayner, S., & Johnston, S. A. (2002). ORF-FINDER: A vector for high-throughput gene identification. *Gene*, *282*(1–2), 33–41.

Roslan, N. D., Tan, C., Ismail, I., & Zainal, Z. (2013). cDNA cloning and expression analysis of the chalcone synthase gene (CHS) from *Polygonum minus*. *Australian Journal of Crop Science*, *7*(6), 777–783.

Roslan, N. D., Yusop, J. M., Baharum, S. N., Othman, R., Mohamed-Hussein, Z. A., Ismail, I., … Zainal, Z. (2012). Flavonoid biosynthesis genes putatively identified in the aromatic plant *Polygonum minus* via expressed sequences tag (EST) analysis. *International Journal of Molecular Sciences*, *13*(3), 2692–2706.

Rosler, J., Krekel, F., Amrhein, N., & Schmid, J. (1997). Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. *Plant physiology*, *113*(1), 175-179.

Roughley, P. J., & Whiting, D. A. (1973). Experiments in the biosynthesis of curcumin. *Journal of the Chemical Society, Perkin Transactions 1*, 2379-2388.

Rubio-Pina, J. A., & Zapata-Perez, O. (2011). Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants. *Electronic Journal of Biotechnology*, *14*(5).

Sadiq, A., Hayat, M. Q., & Mall, S. M. (2014). Qualitative and quantitative determination of secondary metabolites and antioxidant potential of *Eruca sativa*. *Natural Products Chemistry & Research*, *2*(4), 1–7.

Sadowski, J., Gasteiger, J., & Klebe, G. (1994). Comparison of automatic three-dimensional model builders using 639 X-ray structures. *Journal of Chemical Information and Modeling*, *34*(4), 1000–1008.

Salzman, R. A., Fujita, T., Zhu-Salzman, K., Hasegawa, P. M., & Bressan, R. A. (1999). An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Molecular Biology Reporter*, *17*(1), 11-17.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (No. Ed. 2). Cold spring harbor laboratory press.

Sandeep, I., Sanghamitra, N., & Sujata, M. (2015). Differential effect of soil and environment on metabolic expression of tumeric. *Indian Journal of Experimental Biology*, *56*(6), 406–11.

Santhi, R., & Sheeja, T. E. (2013). Deep sequencing identifies candidate miRNAs from turmeric with possible regulatory roles on plant and human genes. In: Sasikumar, B., Dinesh, R., Prasath, D,, Biju, C. N., Srinivasan, V. (eds) National symposium on spices and aromatic crops (SYMSAC VII): Post-Harvest processing of spices and fruit crops. *Indian Society for spices*, Kozhikode, p 210.

Sanwal, S. K., Laxminarayana, K., Yadav, R. K., Rai, N., Yadav, D. S., & Bhuyan, M. (2007). Effect of organic manures on soil fertility, growth, physiology, yield and quality of turmeric. *Indian Journal of Horticulture*, *64*(4), 444-449.

Sasikumar, B. (2005). Genetic resources of Curcuma: diversity, characterization and utilization. *Plant Genetic Resources*, *3*(2), 230–251.

Sasikumar, B., George, J. K., & Zachariah, T. J. (1996). I1SR Prabha and IISR Prathibha- two new high yielding and high quality turmeric (*Curcuma longa* L.) varieties. *Journal of Spices and Aromatic Crops*, *5*, 41–48.

Sato, D. (1948). The karyotype and phylogeny of Zingiberaceae. *The Japanese Journal of Genetics*, 23: 44.

Schenck, C. A., Holland, C. K., Schneider, M. R., Men, Y., Lee, S. G., Jez, J. M., & Maeda, H. A. (2017). Molecular basis of the evolution of alternative tyrosine biosynthetic pathways in plants. *Nature Chemical Biology*, Sep;13(9):1029-1035

Schmidtke, C., Findeiß, S., Sharma, C. M., Kuhfuß, J., Hoffmann, S., Vogel, J., ... & Bonas, U. (2011). Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. *Nucleic Acids Research*, *40*(5), 2020–2031.

Schneider, C., Gordon, O. N., Edwards, R. L., & Luis, P. B. (2015). Degradation of curcumin: From mechanism to biological implications. *Journal of Agricultural and Food Chemistry*, *63*(35), 7606–7614.

Schroder, J. (1997). A family of plant-specific polyketide synthases: Facts and predictions. *Trends in Plant Science*, *2*(10), 373–378.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., … Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, *7*(1), 3.

Sebastiana, M., Figueiredo, A., Monteiro, F., Martins, J., Franco, C., Coelho, A. V., … Ferreira, S. (2013). A possible approach for gel-based proteomic studies in recalcitrant woody plants. *SpringerPlus*, *2*(1), 210.

Shafrin, F., Ferdous, A. S., Sarkar, S. K., Ahmed, R., Amin, A.-, Hossain, K., … Khan, H. (2017). Modification of monolignol biosynthetic pathway in jute: Different gene, different consequence. *Scientific Reports*, *7*(January), 39984.

Shagin, D. A., Rebrikov, D. V., Kozhemyako, V. B., Altshuler, I. M., Shcheglov, A. S., Zhulidov, P. A., … Lukyanov, S. (2002). A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Research*, *12*(12), 1935–1942.

Shah, N.C. (1997) Traditional uses of turmeric (*C. longa* L.) in India. *Journal of Medicinal and Aromatic Plants Science* 19: 948–954.

Shanmugasundaram, K. A., Thangaraj, T., Azhakiamanavalan, R. S., & Ganga, M. (2001). Evaluation and selection of turmeric (*Curcuma longa* L.) genotypes. *Journal of Spices and Aromatic Crops*, *10*(1), 33-36.

Sharma, A. K & Bhattacharyya, N. K. (1959). Cytology of several members of Zingiberaceae. *La Cellule 59*, 197–346.

Sheeja, T. E., Deepa, K., Santhi, R., & Sasikumar, B. (2015). Comparative transcriptome analysis of two species of *Curcuma* contrasting in a high-value compound curcumin: Insights into genetic basis and regulation of biosynthesis. *Plant Molecular Biology Reporter*, 1825–1836.

Shen, H., He, X., Poovaiah, C. R., Wuddineh, W. A., Ma, J., Mann, D. G. J., … Dixon, R. A. (2012). Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist*, *193*(1), 121–136.

Shi, M.-Z., & Xie, D.-Y. (2014). Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Patents on Biotechnology*, *8*(1), 47–60.

Shimada, N., Nakatsuka, T., Nakano, Y., Kakizaki, Y., Abe, Y., Hikage, T., & Nishihara, M. (2009). Identification of gentian cultivars using SCAR markers based on intron-length polymorphisms of flavonoid biosynthetic genes. *Scientia Horticulturae*, *119*(3), 292–296.

Shirley, B. W., Kubasek, W. L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F. M., & Goodman, H. M. (1995). Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *The Plant Journal*. Nov;8(5):659-71.

Shoba, G., Joy, D., Joseph, T., Majeed, M., Rajendran, R., & Srinivas, P. (1998). Influence of piperine on the harmacokinetics of curcumin in animals and human volunteers. *Planta Medica*, *64*(5), 353–356.

Silveira, E. D., Alves-Ferreira, M., Guimarães, L. A., da Silva, F. R., & Carneiro, V. T. D. C. (2009). Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha*. *BMC Plant Biology*, *9*, 84.

Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, *7*(1), 33.

Singh, B., & Ramakrishna, Y. (2014). Indian collections of turmeric (*Curcuma longa* L.): Genetic Variability, inheritance, character association and performance. *Indian Journal of Plant Genetic. Resources, 27*(273), 263–270.

Singh, G., Kumar, S., & Singh, P. (2003). A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. *Plant Molecular Biology Reporter*, *21*(1), 93.

Singh, N., Jain, N., Kumar, R., Jain, A., Singh, N. K., & Rai, V. (2015). A comparative method for protein extraction and 2-D gel electrophoresis from different tissues of *Cajanus cajan*. *Frontiers in Plant Science*, *6*(August), 606.

Sivanandhan, G., Selvaraj, N., Ganapathi, A., & Manickavasagam, M. (2014). Enhanced biosynthesis of withanolides by elicitation and precursor feeding in cell suspension culture of *Withania somnifera* (L.) dunal in shake-flask culture and bioreactor. *PLoS ONE*, *9*(8).

Smarda, P., Bures, P., Horova, L., Leitch, I. J., Mucina, L., Pacini, E., ... & Rotreklová, O. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences*, *111*(39), E4096-E4102.

Smita, S., Rajwanshi, R., Lenka, S. K., Katiyar, A., Chinnusamy, V., & Bansal, K. C. (2013). Comparative analysis of fruit transcriptome in tomato (*Solanum lycopersicum*) genotypes with contrasting lycopene contents. *Plant molecular biology reporter*, *31*(6), 1384-1396.

Song, H. H., Ryu, H. W., Lee, K. J., Jeong, I. Y., Kim, D. S., & Oh, S. R. (2014). Metabolomics investigation of flavonoid synthesis in soybean leaves depending on the growth stage. *Metabolomics*, 1–9.

Srinivasan, V., Thankamani, C. K., Dinesh, R., Kandiannan, K., Zachariah, T. J., Leela, N. K., … Ansha, O. (2016). Nutrient management systems in turmeric: Effects on soil quality, rhizome yield and quality. *Industrial Crops and Products*, *85*, 241–250.

Stahlberg, A., Håkansson, J., Xian, X., Semb, H., & Kubista, M. (2004). Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry*, *50*(3), 509–515.

Strimpakos, A. S., & Sharma, R. A. (2008). Curcumin: preventive and therapeutic properties in laboratory studies and clinical trials. *Antioxid Redox Signal*, *10*(3), 511–545.

Sun, S.-S., Gugger, P. F., Wang, Q.-F., & Chen, J.-M. (2016). Identification of a *R2R3-MYB* gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (*Nelumbo* Adans.). *PeerJ*, *4*, e2369.

Sun, W., Meng, X., Liang, L., Jiang, W., Huang, Y., He, J., … Wang, L. (2015). Molecular and biochemical analysis of chalcone Synthase from freesia hybrid in flavonoid biosynthetic pathway. *PLoS ONE*, *10*(3), 1–18.

Suzuki, Y., Kawazu, T., & Koyama, H. (2004). RNA isolation from siliques, dry seeds, and other tissues of *Arabidopsis thaliana*. *BioTechniques*, *37*(4), 542–544.

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A., & Sugano, S. (1997). Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene*, *200*(1–2), 149–156.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W., & Kubista, M. (2015). How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*, *3*, 9–16.

Syamkumar, S., Lowarence, B., & Sasikumar, B. (2003). Isolation and amplification of DNA from rhizomes of turmeric and ginger. *Plant Molecular Biology Reporter*, *21*(2), 171–171.

Tanaka, Y., Sasaki, N., & Ohmiya, A. (2008). Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant Journal*, *54*(4), 733–749.

Tao, K. E., Dong, C. H., Han, M. A. O., Zhao, Y. Z., Liu, H. Y., & Liu, S. Y. (2011). Construction of a normalized full-length cDNA library of Sesame developing seed by DSN and SMART^{TM}. *Agricultural Sciences in China*, *10*(7), 1004–1009.

Taura, F., Iijima, M., Yamanaka, E., Takahashi, H., Kenmoku, H., Saeki, H., … Morita, H. (2016). A novel class of plant type III polyketide synthase involved in orsellinic acid biosynthesis from *Rhododendron dauricum*. *Frontiers in Plant Science*, *7*(September), 1–15. http://doi.org/10.3389/fpls.2016.01452

Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y., & Morimoto, S. (2009). Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Letters*, *583*(12), 2061–2066.

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., & Nguyen, M. (2010). A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods*, *50*(4), S1–S5.

Thatcher, S. R., Burd, S., Wright, C., Lers, A., & Green, P. J. (2015). Differential expression of miRNAs and their target genes in senescing leaves and siliques: insights from deep sequencing of small RNAs and cleaved target RNAs. *Plant, cell & environment*, *38*(1), 188-200.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

Thomsen, R., & Christensen, M. H. (2006). MolDock: a new technique for high accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*, 3315–3321.

Tian, H., Xu, X., Zhang, F., Wang, Y., Guo, S., Qin, X., & Du, G. (2015). Analysis of *Polygala tenuifolia* transcriptome and description of secondary metabolite biosynthetic pathways by Illumina sequencing. *International Journal of Genomics*, *2015*(August).

Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J. I., Awazuhara, M., … Saito, K. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant Journal*, *42*(2), 218–235.

Torre, S., Tattini, M., Brunetti, C., Guidi, L., Gori, A., Marzano, C., … Sebastiani, F. (2016). *De Novo* assembly and comparative transcriptome analyses of red and green morphs of sweet basil grown in full sunlight. *PLoS ONE*, *11*(8), 1–19.

Tsao, R. (2010). Chemistry and biochemistry of dietary polyphenols. *Nutrients*, *2*(12), 1231–1246.

Upadhyay, S., Phukan, U. J., Mishra, S., & Shukla, R. K. (2014). *De novo* leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in *Asparagus racemosus*. *BMC Genomics*, *15*(1), 746. http://doi.org/10.1186/1471-2164-15-746

Vaganan, M. M., Sarumathi, S., Nandakumar, A., Ravi, I., & Mustaffa, M. M. (2015). Evaluation of different protein extraction methods for banana (*Musa* spp.) root proteome analysis by two-dimensional electrophoresis, *52*(February), 101–106.

Vaidya, K., Ghosh, A., Kumar, V., Chaudhary, S., Srivastava, N., Katudia, K., … Chikara, S. K. (2013). *De novo* transcriptome sequencing in *Trigonella foenum-graecum* L. to identify genes involved in the biosynthesis of diosgenin. *The Plant Genome*, *6*(August), 1–11.

Valasek, M. A., & Repa, J. J. (2005). The power of real-time PCR. *AJP: Advances in Physiology Education*, *29*(3), 151–159.

Valcu, C. M., & Schlink, K. (2006). Efficient extraction of proteins from woody plant samples for two-dimensional electrophoresis. *Proteomics*, *6*(14), 4166–4175.

Van Cutsem, E., Simonart, G., Degand, H., Faber, A. M., Morsomme, P., & Boutry, M. (2011). Gel-based and gel-free proteomic analysis of *Nicotiana tabacum* trichomes identifies proteins involved in secondary metabolism and in the (a)biotic stress response. *Proteomics*, *11*(3), 440–454.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., & Van de Peer, Y. (2009). Unraveling transcriptional control in Arabidopsis using *cis*-regulatory elements and coexpression networks. *Plant Physiology*, *150*(2), 535-546.

Vandesompele, J., De Preter, K., Pattyn, ilip, Poppe, B., Van Roy, N., De Paepe, A., & Speleman, rank. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(711), 34–1.

Velayudhan, K. C., Amalraj, V. A., & Muralidharan, V. K. (1996). The conspectus of the genus Curcuma in India. *Journal of Economic and Taxonomic Botany*, *20*(2), 375-382.

Vogel, A. (1842). Mémoire sur la Curcumine. *Journal de pharmacie et de chimie*, *2*, 20-27.

Vogel, A., & Pelletier, J. (1815). Examen chimique de la racine de Curcuma. *Journal de Pharmacie*, *1*, 289-300.

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Molecular Plant*, *3*(1), 2–20.

Wan, H., Cui, Y., Ding, Y., Mei, J., Dong, H., Zhang, W., … Qian, W. (2017). Time-series analyses of transcriptomes and proteomes reveal molecular networks underlying oil accumulation in Canola. *Frontiers in Plant Science*, *7*(January), 2007.

Wang, G. X., Cao, F. L., & Chen, J. (2006). Progress in researches on the pharmaceutical mechanism and clinical application of *Ginkgo biloba* extract on various kinds of diseases. *Chinese Journal of Integrative Medicine*, *12*(3), 234–239.

Wang, J., Wang, Y., Gu, W., Ni, B., Sun, H., Yu, T., … Shao, Y. (2016). Comparative transcriptome analysis reveals substantial tissue specificity in human aortic valve, *Evolutionary Bioinformatics Online*, *12* (Jul 31), 175-84.

Wang, M., Vannozzi, A., Wang, G., Zhong, Y., Corso, M., Cavallini, E., & Cheng, Z. M. M. (2015). A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation with WRKY proteins. *Frontiers in plant science*, *6*.

Wang, S. (2016). Engineering of polyketide biosynthetic pathways for bioactive molecules, Ph.D Thesis, Utah State University.

Wang, T., Zhang, N., & Du, L. (2005a). Isolation of RNA of high quality and yield from *Ginkgo biloba* leaves. *Biotechnology Letters*, *27*(9), 629–633.

Wang, X., Zhao, X., Zhu, J., & Wu, W. (2005b). Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (Oryza sativa L.). *DNA research*, *12*(6), 417-427.

Wang, Y. J., Pan, M. H., Cheng, A. L., Lin, L. I., Ho, Y. S., Hsieh, C. Y., & Lin, J. K. (1997). Stability of curcumin in buffer solutions and characterization of its degradation products. *Journal of Pharmaceutical and Biomedical Analysis*, *15*(12), 1867–1876.

Wang, Y., Hu, W. Y., & Wang, M. Z. (1999). HPLC determination of three curcuminoid constituents in rhizoma Curcumae. *Acta Pharmaceutica Sinica*, *34*, 470-473.

Weber, A. P. M. (2015). Discovering new biology through sequencing of RNA. *Plant Physiology*, *169*(3), 1524–31.

Wei, H., Chen, X., Zong, X., Shu, H., Gao, D., & Liu, Q. (2015). Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). *PLoS ONE*, *10*(3), 1–20.

Wei, Y. Z., Hu, F. C., Hu, G. B., Li, X. J., Huang, X. M., & Wang, H. C. (2011). Differential expression of anthocyanin biosynthetic genes in relation to anthocyanin accumulation in the pericarp of *Litchi chinensis* sonn. *PLoS ONE*, *6*(4).

Wentzell, A. M., & Kliebenstein, D. J. (2008). Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. *Plant physiology*, *147*(1), 415-428.

Wiemann, S., Mehrle, A., Bechtel, S., Wellenreuther, R., Pepperkok, R., & Poustka, A. (2003). cDNAs for functional genomics and proteomics: The German Consortium. *Comptes Rendus - Biologies*, *326*(10–11), 1003–1009.

Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA translation by 5′- and 3′-UTR-binding factors. *Trends in Biochemical Sciences*, *28*(4), 182–188.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., … Hochstrasser, D. F. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology*, *14*(1), 61–5.

Wu, J., Xiao, J., Wang, L., Zhong, J., Yin, H., Wu, S., ... & Yu, J. (2013). Systematic analysis of intron size and abundance parameters in diverse lineages. *Science China. Life Sciences*, *56*(10), 968.

Wu, Q., Wu, J., Li, S.-S., Zhang, H.-J., Feng, C.-Y., Yin, D.-D., … Wang, L.-S. (2016). Transcriptome sequencing and metabolite analysis for revealing the blue flower formation in waterlily. *BMC Genomics*, *17*(1), 897.

Xiang, L. L., Liu, X. F., Li, X., Yin, X. R., Grierson, D., Li, F., & Chen, K. S. (2015). A Novel bHLH Transcription Factor Involved in Regulating Anthocyanin Biosynthesis in Chrysanthemums (*Chrysanthemum morifolium* Ramat.). *PloS one*, *10*(11), e0143892.

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J. S., Chakrabarty, R., … Sensen, C. W. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology*, *166*(3), 122–134.

Xie, F., Xiao, P., Chen, D., Xu, L., & Zhang, B. (2012). miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology*, *80*(1), 75–84.

Xie, R., Zheng, L., He, S., Zheng, Y., Yi, S., & Deng, L. (2011). Anthocyanin biosynthesis in fruit tree crops: genes and their regulation. *African Journal of Biotechnology*, *10*(86), 19890-19897.

Xie, R., Zheng, L., He, S., Zheng, Y., Yi, S., & Deng, L. (2014). Anthocyanin biosynthesis in fruit tree crops: Genes and their regulation. *African Journal of Biotechnology*, *10*(86), 19890–19897.

Xie, Z., Ma, X., & Gang, D. R. (2009). Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism. *Journal of Experimental Botany*, *60*(1), 87–97.

Xu, C., Jiao, C., Zheng, Y., Sun, H., Liu, W., Cai, X., … Wang, Q. (2015). *De novo* and comparative transcriptome analysis of cultivated and wild spinach. *Scientific Reports*, *5*, 17706.

Yamamoto, Y. Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., … Abe, T. (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics*, *8*(8), 67.

Yang, C. Q., Fang, X., Wu, X. M., Mao, Y. B., Wang, L. J., & Chen, X. Y. (2012). Transcriptional regulation of plant secondary metabolism. *Journal of integrative plant biology*, *54*(10), 703-712.

Yao, L. M., Wang, B., Cheng, L. J., & Wu, T. L. (2013). Identification of key drought stress-related genes in the hyacinth bean. *PLoS ONE*, *8*(3).

Yuan, J. S., Reed, A., Chen, F., & Stewart, C. N. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics*, *7*, 85.

Zachariah, T. J., Sasikumar, B., & Nirmal Babu, K. (1999). Variation for quality components in ginger and turmeric and their interaction with environments. In *Proceedings of National Seminar on Biodiversity, Conservation and Utilisation of spices, Medicinal and Aromatic Plants. Calicut: IISR* (pp. 116-120).

Zandalinas, S. I., Sales, C., Beltrán, J., Gómez-Cadenas, A., & Arbona, V. (2016). Activation of secondary metabolism in citrus plants is associated to sensitivity to combined drought and high temperatures. *Frontiers in plant science*, *7*.

Zhai, R., Wang, Z., Zhang, S., Meng, G., Song, L., Wang, Z., ... & Xu, L. (2015). Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.). *Journal of experimental botany*, *67*(5), 1275-1284.

Zhan, C., Li, X., Zhao, Z., Yang, T., Wang, X., Luo, B., … Hu, X. (2016). Comprehensive analysis of the triterpenoid saponins biosynthetic pathway in *Anemone flaccida* by transcriptome and proteome profiling. *Frontiers in Plant Science*, *7*(July), 1–12.

Zhang, L., Wang, L., Yang, Y., Cui, J., Chang, F., Wang, Y., & Ma, H. (2015). Analysis of *Arabidopsis* floral transcriptome: detection of new florally expressed genes and expansion of Brassicaceae-specific gene families. *Frontiers in Plant Science*, *5*(January), 1–11.

Zhao, N., Wang, G., Norris, A., Chen, X., & Chen, F. (2013). Studying plant secondary metabolism in the age of genomics. *Critical Reviews in Plant Sciences*, *32*(6), 369–382.

Zhao, X., Si, J., Miao, Y., Peng, Y., Wang, L., & Cai, X. (2014). Comparative proteomics of *Euphorbia kansui* Liou milky sap at two different developmental stages. *Plant Physiology and Biochemistry*, *79*, 60–65.

Zhao, X., Yang, L., Zheng, Y., Xu, Z., & Wu, W. (2009). Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). *Journal of Genetics and Genomics*, *36*(7), 435–442.

Zhao, Y., Liu, T., Luo, J., Zhang, Q., Xu, S., Han, C., … Kong, L. (2015). Integration of a decrescent transcriptome and metabolomics dataset of *Peucedanum praeruptorum* to investigate the CYP450 and MDR genes involved in coumarins biosynthesis and transport. *Frontiers in Plant Science*, *6*(December), 996.

Zhou, A., Ma, H., Liu, E., Jiang, T., Feng, S., Gong, S., & Wang, J. (2017). Transcriptome sequencing of *Dianthus spiculifolius* and analysis of the genes involved in responses to combined cold and drought stress. *International Journal of Molecular Sciences*, *18*(4).

Zhou, L., Wang, Y., & Peng, Z. (2011). Molecular characterization and expression analysis of chalcone synthase gene during flower development in tree peony (*Paeonia suffruticosa*). *African Journal of Biotechnology*, *10*(8), 1275–1284.

Zhou, W. Z., Zhang, Y. M., Lu, J. Y., & Li, J. F. (2012). Construction and evaluation of normalized cDNA libraries enriched with full-length sequences for rapid discovery of new genes from sisal (*Agave sisalana* perr.) different developmental stages. *International Journal of Molecular Sciences*, *13*(10), 13150–13168.

Zhu, Q., Xie, X., Lin, H., Sui, S., Shen, R., Yang, Z., … Liu, Y. G. (2015). Isolation and functional characterization of a phenylalanine ammonia-lyase gene (SsPAL1) from coleus (*Solenostemon scutellarioides* (L.) Codd). *Molecules*, *20*(9), 16833–16851.

Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., & Siebert, P. D. (2001). Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *Biotechniques*, *30*(4), 892-897.

Zhulidov, P. A., Bogdanova, E. A., Shcheglov, A. S., Vagner, L. L., Khaspekov, G. L., Kozhemyako, V. B., … Shagin, D. a. (2004). Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research*, *32*(3), e37.

Zuiter, A. S., Sawwan, J., & Al Abdallat, A. (2012). Designing universal primers for the isolation of DNA sequences encoding proanthocyanidins biosynthetic enzymes in *Crataegus aronia*. *BMC Research Notes*, *5*(1), 427.

Zulak, K. G., Khan, M. F., Alcantara, J., Schriemer, D. C., & Facchini, P. J. (2009). Plant defense responses in opium poppy cell cultures revealed by liquid chromatography-Tandem mass spectrometry proteomics. *Molecular & Cellular Proteomics*, *8*(1), 86–98.

Appendix...

Appendix 1. List of primers used in the study

Name	Forward primer (5'→3')	Reverse primer (5'→3')
Primers to check the efficiency of cDNA		
ACTIN	GGTTGGTATGGGT C	TGCAATCCACATCTGTTGGAAG
Sequence of adaptor primer used for ds cDNA synthesis		
PCR primer M1	AAGCAGTGGTATCAACGCAGAGT	
Primers to amplify the target gene from the recombinant clone		
M13	GTAAAACGACGGCCAG	CAGGAAACAGCTATGAC
Primers for qPCR analysis		
EF1α	GCTGACTGTGCTGTTCTCATTAT	CTCGTGTCTGTCCATCCTTTGAA
UBIQUITIN	GCACTCTCGCTGACTACAAC	GGCTTGGTGTAGGTCTTCTTC
18S rRNA	CCTTCCTCTAAATGATAAGGTTCAATG	GATTGAATGGTCCGGTGAAGTGTT
GAPDH	AACTGTAGCCCCACTCATTG	GCATCTTAGGGTATGTGGAGG
ACTIN	CAACAGCAGAACGGGAAATTG	CATAATCAAGGGCGACATATGC
TUBULIN	GGCAGAGATCAGATGGTTCAG	TGGACAATGAAGCACTCTACG
PAL	GTACAGCGGGTACGACCTA	GCTTTCGAGGAAGAGCTCAA
C4H	TTACTTGCAGGCGGTGATC	AGGCGTTGACCAGTATCTTG
4CL	GTGGAGTCTTCAGGGAAGAGAG	GGCGATTGTCGATGTGAATG
C3H	GATGGTCACCTTCATGCATACT	GGAAGAAGCTGCAGGAGTAATAG
HCT	TTCATCGACAACCCCAAGAC	ATCGGAGACATTGGGAAGC
COMT	TGCCGGAGGGTCAAATG	ACCAGAAACTCAACTGAAAGGG
DCS	GTGCTGTTCATCCTGGACGAG	CAACAGCACGCCCCAGTCGA
CURS1	TCAGCTCATCCATCACGAAGTACAC	CATCATTGACGCCATCGAAGC
CURS2	TGTTGCCGAACTCGGAGAAGAC	TCGGGATCAGGACTGGAACAAC
CURS3	CCCATTCCTTGATCGCCTTTTCC	TGGAGCCCTCCTTCGACGACC
ClPKS-a	TGTCGGAGATCACCCACTTG	CGGAGAAGGAGAGACCGAGA
ClPKS-b	TTCCTCACGTCGTCCATCAC	AGGCACGTCTTCAGCGAGTT
ClPKS-c	CCTCCATGGCCTCCAACTAC	CGTAGAGGTTGGGAGGGTTG
ClPKS-d	CACGTCTCCAAGTTGTTGGC	CCGATCTTCGAGATCATGGC
ClPKS-e	AGGCACTGTTGGATGTTCG	TGAGGGAGATCGGGTTGAC
ClPKS-f	GAGGCCACTTCGTAGATCG	ACCACTTCGAGAACCTCATTG
WRKY	GGCTATCAGGCAGGTTCAGG	TTTGAGGAACCAAGGGAGGA
bHLH	CGTACAAGCGAAAAGCAGAAG	ACATTGACATCGATAGGGCC
MYB1	GATCCTAGCAGAGCATGTCAG	ATGTTTCCCCTCTTGATGTCC
MYB2	CAAGGGAGTTTGGACCAAGG	AGCAAGCCAGCAGCTTTAGG
Primers for amplifying full length coding region		
DCS-K	CCGAATTCCATATGGAAGCGAACGGCTACCG	CGCGGATCCCTAGTTCAGTCTGCAACTAT
CURS1-K	CCGAATTCCATATGGCCAACCTCCACGCGTT	CGCGGATCCCTACAGTGGCATACTGCGCA
CURS3-K	CCGAATTCCATATGGGCAGCCTGCAGGCGAT	CGCGGATCCCTACGGTAATGGTACACTGC
CURS3-FL	TAGCTAGCTGCAATTCGTTAAT	AGTGTACAGGTGATCGCA
ClPKS11-FL	CAGCAACCAGTTTCGCTTTC	CGCCGGAATAAGCTACAGAG
Primers for inverse PCR reaction		
CURS3-IP	GGCTGGACCTTCCACTTCTT	TTCGCCAGCTTTGGTATCTC
ClPKS11-IPa	TCGTCGAAGGAAGTGTCCAT	CGATATGCCTGGAGCTGATT
ClPKS11-IPb	GCCAAACTTCTCGGTCTCTC	GTCCATGTACGAGCACATCC
Primers for pre-mRNA analysis		
*ClPKS11*exon-1	ACCGGTGCCGACCACAA	
*ClPKS11*exon-2	CTTCAGCAGCTCCTCCG	
*ClPKS11*intron	CGCAACACGATTAGCGTAAC	
Primers used for amplifying upstream region of *ClPKS11*		
PMTR-IPa	TCGTCGAAGGAAGTGTCCAT	CGATATGCCTGGAGCTGATT
PMTR-IPb	CTGGGCCTTGAGCCCTTT	GCCAAACTTCTCGGTCTCTC
PMTR	TTCTTCCTTCTTACACGGTCAA	CTGGGCCTTGAGCCCTTT

Publications and awards

Publications and awards

Research articles published

1. **Deepa, K**., Sheeja, T. E., Rosana, O. B., Srinivasan, V., Krishnamurthy, K. S., & Sasikumar, B. (2017). Highly conserved sequence of *ClPKS11* encodes a novel polyketide synthase involved in curcumin biosynthesis in turmeric (*Curcuma longa* L.). *Industrial Crops and Products*, *97*, 229-241. **NAAS rating-9.45; Impact Factor: 3.181**
2. Sheeja, T. E., **Deepa, K**., Santhi, R., & Sasikumar, B. (2015). Comparative transcriptome analysis of two species of *Curcuma* contrasting in a high-value compound curcumin: Insights into genetic basis and regulation of biosynthesis. *Plant molecular biology reporter*, *33*(6), 1825-1836. **NAAS rating-8.30; Impact Factor:1.932**
3. **Deepa, K**., Sheeja, T. E., Santhi, R., Sasikumar, B., Cyriac, A., Deepesh, P. V., & Prasath, D. (2014). A simple and efficient protocol for isolation of high quality functional RNA from different tissues of turmeric (*Curcuma longa* L.). *Physiology and Molecular Biology of Plants*, *20*(2), 263-271. **NAAS rating-7.35; Impact Factor: 0.883**

Presentation in Seminars/Symposia

1. **Deepa K**, Sheeja TE, Rosana OB and Sasikumar B. Highly conserved sequence of *ClPKS11* is a novel gene involved in *de novo* curcumin biosynthesis in turmeric (*Curcuma longa* L.), SYMSAC VIII during 16-18 December 2015, Coimbatore
2. Sheeja TE, **Deepa K**, Santhi R and Sasikumar B. Comparative transcriptome analysis of *Curcuma longa* (turmeric) and *Curcuma aromatica* (wild turmeric) provide insights into genetic basis of curcuminoid biosynthesis and its regulation, 2014 NextGen Genomics & Bioinformatics Technologies (NGBT) Conference during 17-19 November 2014, NIMHANS, Bangalore, India.
3. **Deepa K**, Sheeja TE and Sasikumar B. Construction of a normalized full length cDNA library from turmeric for rapid gene discovery, SYMSAC VII during 27-29 November 2013, Madikkeri

Awards

1. **Dr Alwar Memorial Prize**-2017 for Best Research Fellow, ICAR-IISR
2. **Best Poster award** : **Deepa K**, Sheeja TE, Rosana OB and Sasikumar B. Highly conserved sequence of *ClPKS11* is a novel gene involved in *de novo* curcumin biosynthesis in turmeric (*Curcuma longa* L.), SYMSAC VIII during 16-18 December 2015, Coimbatore