Diversity and Characterization of Curcumin Biosynthetic Genes and Transcription Factors from *Curcuma* Spp.

A thesis submitted to University of Calicut

for the degree of Doctor of Philosophy

(Biotechnology)

By

Prashina Mol P

Under the guidance of

Dr. Sheeja TE

Division of Crop Improvement and Biotechnology

ICAR-Indian Institute of Spices Research

Kozhikode, 673012, Kerala, India

December 2021

भाकृअनुप - भारतीय मसाला फसल अनुसंधान संस्थान
ICAR - INDIAN INSTITUTE OF SPICES RESEARCH
(भारतीय कृषि अनुसंधान परिषद *Indian Council of Agricultural Research*)
पी. बी संख्याः *Post Bag No: 1701*, मेरिकुन्नु पोस्ट *Marikunnu Post*,
कोषिक्कोड *Kozhikode-673012*, केरल, *Kerala*, भारत *India*
(ISO 9001: 2008 Certified Institute)

CERTIFICATE

This is to certify that the thesis entitled "**Diversity and Characterization of Curcumin Biosynthetic Genes and Transcription Factors from *Curcuma* spp.**" submitted by Mrs. Prashina Mol P., 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 30-03-2016 to 29-12-2021. This thesis has not been submitted for the award of any other degree or diploma of this or any other University.

(Dr. Sheeja TE)

Head of the institute **Signature of the Guide**

Place : Kozhikode

Date :

पी ए बी एक्स PABX: 0495-2731410/2731753/2731345 निदेशक का कार्यालय Director's Office: 0495-2730294 परियोजना समन्वयक Project Coordinator: 0495-2731794, एटिक ATIC: 0495 - 2730704. आई आई एस आर प्रयोगिक प्रक्षेत्र, पेरुवण्णामुषि ISSR Experimental Farm, Peruvannamuzhi: 0496 2249371 कृषि विज्ञान केन्द्र Krishi Vigyan Kendra, पेरुवण्णामुषि Peruvannamuzhi: 0496-2666041, तार Grams: RESEARCH CALICUT. फैक्स Fax: 0091-495-2731187 ई-मेल Email: mail@spices.res.in वेबसाईट Website: www.spices.res.in

DECLARATION

I hereby declare that the thesis entitled "**Diversity and Characterization of Curcumin Biosynthetic Genes and Transcription Factors from *Curcuma* spp.**" 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.

(Prashina Mol P)

Place : Kozhikode

Date :

ACKNOWLEDGEMENT

First and foremost, I praise **God**, the Almighty for his showers of blessings throughout my research which helped me to complete my research work successfully.

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

My sincere thanks to Dr. J. Rema, Director in-charge, Dr. Nirmal Babu, Dr. M. Anandaraj former Directors, Indian Institute of Spices Research for permitting and extending me the facilities to carry out this work at the Indian Institute of Spices Research, Kozhikode.

I would like to express my heartfelt thanks to Dr. J. Rema, Head of Division and B. Sasikumar, former Head of Division for their continuous support during my research work.

I express my sincere thanks to UGC, New Delhi for UGC-JRF fellowship and I thank University of Calicut for all the necessary help and support right from the registration to completion of this Ph. D.

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

I would like to thank Dr. Santhosh J. Eapen, Head and all the Scientists of Division of Crop Protection and Dr. C Thankamani, Head and all the Scientists of Division of Crop Production and Post-harvest Technology.

I thank Dr. V Srinivasan, Dr K.S. Krishnamurthy, Dr. R.R. Nair and Dr. D. Prasath for providing samples and assisting in field experiments

I thank Dr. Sona Charles for helping me in bioinformatic work

I thank Shri. Krishnadas farm officer, Shri. Binoy, Shri. Vijesh Kumar, Shri. Sudakaran, Shri. Jayaprakash, Shri. Sivadas, and all the technical staff and staff of administration and accounts section, especially Shri. Sayed Mohammed, Shri. Sundaran,

It is a great pleasure to thank all my friends Aparna RS, Alka, Aneesha Babu, Aslam, Aswathi K. N. Deepthi, Divya, Dona, Fayad Mohammed, Giridhari, Hemesh k Govindan, Krishnapriya, Megha, Merlin, Nahim, Nazmin, Neenu, Prameela, Raghav, Raguveer, Revathi, Rugma, Shelvy S, Snigdha M, Sreena, Subila, Theertha, Vidya V, Vijayasanthi, Vishnu and Zumila for their help and encouragement. I thank my beloved seniors Mrs. Deepa K, Mrs. Santhi R Mrs. Neema Malik, Mrs. Manju, Mrs. Swetha, Revathy for their love and support.

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

Last but not least, I would like to thank my parents, brothers, grandparents, and all my family members for providing me unfailing support and continuous encouragement throughout my years of study and through the process of writing the thesis.

I am extremely thankful to my husband for the understanding and patience that he has shown to me. His boundless love, co-operation and moral support have helped me to achieve my goal. I express my sincere thanks to my husband and my in laws for the support and encouragement. Finally, my thanks go to all the people who have supported me to complete the research work.

Prashina Mol P

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

µA	Microampere
µg	Microgram
µl	Microlitre
µM	Micromolar
°C	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
$CaCl_2$	Calcium chloride
Cbe-F	Coimbatore-field
cDNA	Complementary DNA
CHS	Chalcone synthase
ClPKS	*Curcuma longa* polyketide synthase
ClPAL2	*Curcuma longa* phenylalanine ammonia lyase 2
ClOMT2	*Curcuma longa* O-methyltransferase 2
CLP	*Curcuma longa* prathibha
CLA	*Curcuma longa* Acc. 200
CA	*Curcuma aromatica*
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	Ethylenediamine tetraacetic 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
$MgCl_2$	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
PPP	Phenylpropanoid pathway
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 guide and my family

BACKGROUND INFORMATION

Genus *Curcuma* (family Zingiberaceae) with approximately 120 species (Sharifi-Rad et al., 2020) has a rich history of medicinal use (Akarchariya et al., 2017; Dosoky and Setzer, 2018). Among the *Curcuma* species, *Curcuma longa* Linn. (Syn: *C. domestica* Val.) or turmeric is the most widely recognized primeval spice. Turmeric, a native of tropical South Asia is primarily cultivated in Asian, Australian and Western African countries (Ewon and Bhagya, 2019), is a perennial rhizomatous herb that grows up to 1 m with a short stem and is cultivated vegetatively as an annual crop for rhizomes. The leafy shoot dies back during the dry season. Turmeric needs an optimum temperature between 20-30°C, as well as a significant amount of annual rainfall to flourish (Prasad and Aggarwal, 2011). Indian turmeric is considered to be best in the world and India is a major supplier of turmeric with more than 60 percent share in turmeric trade (Angles et al., 2011). Rhizomes are the extensively used medicinal part of the turmeric plant (Bhowmik et al., 2009) and turmeric has been used as an ingredient in traditional herbal system of medicines of our country (Muthusamy, 2011).

Turmeric has protein (6.3%), fat (5.1%), minerals (3.5%), carbohydrates (69.4%), and moisture (13.1%) (Chattopadhyay et al., 2004b) with rhizomes containing 3-15% curcuminoids (Lee et al., 2016). Curcuminoid is a complex polyphenol constituted of curcumin (52–63%), demethoxycurcumin (19–27%) and bisdemethoxycurcumin (18–28%) (Hadi et al., 2018). Although curcumin has been reported from different *Curcuma* spp. like *C. xanthorrhiza* (1.5%), *C. angustifolia* (0.2%), *C. zedoaria* (2.0%) it is observed that turmeric or *C. longa* is the main source of curcumin (Lee et al., 2016). Curcumin possess multiple biological activities like antioxidant (Jakubczyk et al., 2020;; Tanvir et al., 2017), anti-inflammatory (Basnet and Skalko-Basnet, 2011; Chainani-Wu, 2003), anti-angiogenic functions (Shakeri et al., 2018), wound healing properties (Akbik et al., 2014) and anti-cancer effects (Pongrakhananon and Rojanasakul, 2011). It is widely used for the treatment of several diseases (Rahmani et al., 2018) in modern medicine too. Curcumin has numerous action mechanisms like antiviral, anti-inflammatory, anticoagulant, antiplatelet, and cytoprotective. Due to these properties and many other biological effects curcumin is a promising target in the adjuvant treatment for COVID-19 (Rattis et al., 2021; Rocha and de Assis, 2020). As the global scenario is now changing towards the use of non-toxic plant products having traditional medicinal use, development of modern drugs from turmeric for the control

of various diseases is gaining momentum (Atanasov et al., 2021). Pharmaceutical industries always prefer high curcumin content varieties. However, curcumin is reported to vary with different varieties of *C. longa* (Guddadarangavvanahally et al., 2002) causing instability in yield of curcumin. This is one of the main concerns of spices industry, as genotypes perform differently across environments (Anandaraj et al., 2014). The main factor governing curcumin biosynthesis is still unknown, with only a rudimentary information available based on biochemical studies.

Curcumin being an oil-soluble compound, is practically insoluble in water at acidic and neutral pH at room temperature (Stohs et al., 2020). Ingesting curcumin as such does not lead to the associated health benefits due to its poor bioavailability, which appears to be primarily because of incomplete absorption, fast metabolism and elimination from the system (Hewlings and Kalman, 2017). A large number of approaches have been utilized to increase the solubility and hence the bioavailability of curcumin. Bioavailability of curcumin is low and for improving it several strategies like incorporating it in a curcumin-piperine complex, delivering in the form of nanoparticles and cyclodextrin inclusions, liposomes and phospholipid complexes (Henrotin et al., 2013). Other strategies include inhibition of curcumin metabolism with adjuvants as well as various oral delivery systems (Cas and Ghidoni, 2019). Micro and nano formulations enhance absorption of curcumin thereby improving bioavailability (Stohs et al., 2020). In spite of development of an array of such formulations and products, a very few of them have reached a commercial level (Jamwal, 2018).

In spite of its global and commercial importance, studies related to curcumin biosynthesis and gene regulation are very scarce. Identification of genes and regulatory proteins involved in curcumin biosynthesis is a basic pre-requisite for improvement of this trait through genetic manipulation. Moreover biosynthetic pathways is the subject of much interest due to the pharmaceutically important properties of curcumin (Katsuyama et al., 2009a). A basic frame work of biosynthesis of curcumin is available through radio tracer studies conducted decades ago (Roughley and Whiting, 1973). Through enzyme kinetic studies, Ramirez-Ahumada et al., 2006 reported that curcumin is synthesized through the Phenylpropanoid pathway (PPP). In many model and non-model plants PPP is well characterized. Study of phenylpropanoid derivatives and other secondary metabolites involved in many plant species revealed the importance of such compounds for the plant survival (Fraser and Chapple, 2011). Phenylpropanoid

metabolism connect the primary and secondary metabolism, where phenylalanine, end product of shikimate pathway gets converted to p-coumaroyl CoA by the action of three enzymes, phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H) 3 and 4-coumarate CoA: ligase (4CL).

Most of the genes involved in PPP exist as multigene families and multiple isoforms are reported for genes in PPP (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, flavonoids *etc*. (Baxter and Stewart, 2013). Type III polyketide synthases are involved in the production of diarylheptanoids like curcumin (Schröder, 1997). Katsuyama et al. (2009 a, b) identified four polyketide synthases namely diketide CoA synthase (DCS) and curcumin synthases (CURS 1-3) from turmeric. A novel type III polyketide synthase named ClPKS10 proposed to have a role in the curcuminoid scaffold biosynthetic pathway was isolated and characterized (Resmi and Soniya, 2012). Sheeja et al., 2015 identified several pathway genes showing differential expression with respect to curcumin through comparative transcriptome analysis of cultivated turmeric (curcumin 6.5%) and a wild species *C. aromatica* (curcumin 0%). Transcriptome and qRT-PCR-based coexpression studies helped to identify a bonafide candidate gene *ClPKS11*, downstream of the curcumin biosynthesis pathway (Deepa et al. 2017, Deepa, 2018).

One of the major mechanisms regulating secondary metabolite production in plant cells is the controlled transcription of biosynthetic genes (Vom Endt et al., 2002). MYB transcription factors are the major regulators of phenylpropanoid metabolism in plants (Zhang et al., 2016). *MYB* related transcription factors of 1R class, which positively regulate curcumin biosynthesis was identified through comparative transcriptome analysis from turmeric (Aparna et al., 2021). Other transcription factors such as WRKY (Lloyd et al., 2017), bZIP (An et al., 2017), MADS (Jaakola et al., 2010) *etc*. were also reported to be involved in PPP. However, some of the recent studies indicate the involvement of either MYB or MYB-bHLH-WDR complex in regulation of PPP (Li, 2014; Lloyd et al., 2017; Nemesio-Gorriz et al., 2017; 2014; Zhao et al., 2019). A couple of bHLH and a single WD40 transcription factor involved in the biosynthesis of curcumin was identified through comparative transcriptome analysis (Mol et al., 2021).

Downstream genes of curcumin pathway have been characterised partially. However, genes involved in the upstream of the pathway remain largely uncharacterized. Moreover, detailed studies involving identification of candidate isoforms and rate limiting enzyme, co-expression and co-regulation network analysis and functional analysis and validation are yet to be carried out. Next generation sequencing (NGS) techniques allows holistic profiling of RNA expression (Ozsolak and Milos, 2011; Wang et al., 2009) and is used to detect novel genes from non-model species for identifying and characterizing secondary metabolism genes and pathways (Reddy et al., 2015; Scossa et al., 2018). RNA sequencing (RNA-seq) technique which uses NGS technology, provides whole-transcriptome expression profiles of selected plant tissues or cells, thereby permitting the integrated analysis of transcriptomes, proteomes and the metabolomes in any plant species (Reddy et al., 2015; Zhao et al., 2019). Comparative transcriptome analysis is a technique used to mine differentially expressed genes by comparing two or more transcriptomes. Comparative transcriptome analysis is now widely used to identify candidate genes and transcription factors involved in plant secondary metabolism (Foong et al., 2020; Jiang et al., 2020; Liu et al., 2015; Soltani et al., 2018). In systems biology, a logically conceivable assumption is that a set of genes involved in a particular biological process (more practically in a secondary metabolic pathway) are co-regulated and thus co-expressed under the control of a shared regulatory system. This is a typical example of the so-called 'guilt-by-association' principle. Based on this, gene co-expression network is widely utilized to identify genes involved in a particular secondary metabolism in plants using guide-gene (*i.e.*, bait gene) or non-targeted protocols (Aoki et al., 2007).

To find physiological roles of genes, the structural characterization of genes by full length cDNA amplification, sequence analysis and molecular docking is compulsory (Bhat et al., 2013). Rapid amplification of cDNA ends (RACE) is a technique used to obtain the full-length sequences of RNA transcripts (Frohman, 1994). RACE can provide the sequence of an RNA transcript from a short, known sequence within the transcript all the way to the 5' end (5' RACE-PCR) or 3' end (3' RACE-PCR) of the RNA.

The advancement of omics has made it possible to investigate the genome wide exploration of all kinds of biological processes at the molecular level. Almost every field of plant biology has been analysed at the genomic, transcriptomic and proteomic

level (Manzoni et al., 2018). High-throughput sequencing is a powerful tool for identifying genes and transcription factors involved in the biosynthetic pathway. Many high-throughput transcriptome data sets have identified many transcriptional changes to date (Diray-Arce et al., 2015). However, mRNA levels and protein levels cannot often be correlated because of post-translational modifications (Maier et al., 2009). Therefore, the use of proteome analyses, which can identify proteins involved in biosynthesis pathway, has become an important strategy. Widely expanded proteomics approaches have been integrated with bioinformatics advances, generating favourable outcomes may help for the elucidation of complex mechanisms involved in curcumin biosynthesis. Proteomics was developed in the early 1990s to allow proteins expressed by cells and tissues to be systematically studied (Arrell et al., 2001; Celis et al., 1999). Major interest in this research relies on the discovery of the genes and enzymes involved in the biosynthesis of bioactive and high-value natural products to bridge in the still many knowledge gaps in their respective pathways. Identification and characterization of regulatory molecules like transcription factors involved in modulating secondary metabolite pathways is a frontier research avenue (Yang et al., 2012).

Proteomics in spite of being the latest NextGen strategy with immense potential, is being put to a very limited use. Proteomics research was originally based on 2-DE protein separation coupled to MS analysis of spots (I^{st} generation). Moving then to LC-based shotgun strategies (2^{nd} generation), and lately to quantification approaches including label and label-free variants (3^{rd} generation). Selected or multiple reaction monitoring (SRM, MRM) hypothesis-driven approaches constitute the last, fourth generation. Proteomics can be divided into subareas including subcellular, comparative, post-translational, interactomics, proteinomics and translational. Most works on plant proteomics so far published belong to the first three categories. Depending on the quantification strategy used (MS-based quantification), we may refer to relative quantitation based on label (DIGE, ICAT, iTRAQ, SILAC) and label-free (peak area or spectral counting) strategies (Jorrin-Novo, 2014).

Shotgun proteomics methods involving isotope labelling of proteins have been developed during the last decade which overcome some of the difficulties associated with quantification using gel-based approaches (Wu et al., 2006). Shotgun proteomics approaches enable identification of proteins that are up-regulated or down-regulated

under specific conditions and this can be studied in different cell and tissue lysates. One strategy, called SILAC (stable isotope labeling by amino acids in cell culture) involves metabolic incorporation of specific amino acids into proteins (Ong et al., 2002). Major drawback is that it relies on endogenous labelling of cell lines (R. and E., 2012) another strategy involves Isotope Coded Affinity Tags (ICAT). It is a cysteine specific, protein-based labelling strategy designed to compare two different sample states (Gygi et al., 1999). Major drawback is that the peptides lacking cysteine residues will not be labelled, so many important peptides, including those with post-translational modifications (PTMs) will be discarded (R. and E., 2012). Isobaric tagging strategies overcome some of the major limitations of isotope tagging. One such method was developed by Applied Biosystems and is called iTRAQ: isobaric tags for relative and absolute quantification (Ross et al., 2004). Isobaric tags for relative and absolute quantification (iTRAQ) make it possible to both identify and quantify proteins simultaneously (R. and E., 2012). The iTRAQ tags react with all primary amines of peptides, which means that all peptides are labelled and information about their post-translational modifications are retained. The isobaric nature of the tags also means that the same peptide from each of the samples being compared appears as a single peak in the mass spectrum (Luo and Zhao et al., 2012). There are very few reports available that have employed iTRAQ for mining and validation of secondary metabolite pathways in rice and *Arabidopsis* (Chen et al., 2016; Wang et al., 2018) and no studies have yet been done in turmeric.

Immense quantity of information being generated in the study from the RNA-seq and protein sequencing, and key genes and regulatory molecules for curcumin biosynthesis being identified, it is prudential to utilise the information for overcoming allied technological backstops too. Identification of curcumin status at early developmental stages emerged as a pre-requisite in breeding programmes for screening for high curcumin lines. The prevalent methods are time consuming and can be done only at harvest. Several colleagues have posed this issue and an attempt was made by us to identify a biomarker for addressing this. Molecular markers, such as transcripts or metabolites, are advantageous as they integrate several different genes and environmental effects. In addition, polymorphic DNA markers may be used for the same purpose too. However, while this is routinely performed in diploid crops, the use of such markers in polyploids remain to be challenging. A simple qRT-PCR technique

based on correlation of gene expression with metabolite content was developed that enables early detection of curcumin status around 120 days after planting (DAP). The basic principle is that higher the curcumin, higher the gene expression and lower the Ct value. The method was confirmed using three key pathway genes that are positive regulators of curcumin. The technique may be exploited for fast, easy and economical way for screening for curcumin and is the first of its kind developed in any spice crop. Similar work involving 31 potato cultivars helped in identification of a set of highly predictive markers for superior drought tolerance (Haas et al., 2020; Sprenger et al., 2018).

AIM AND OBJECTIVES

Even though curcumin is a very important biologically active compound only limited studies have been conducted to explore the molecular mechanisms involved in regulation of biosynthesis. Recent reports revealed that curcumin content is highly dependent on environmental and growth conditions. The current technological advancements like transcriptomics and proteomics are a totally unexplored area of research and a wealth of information can be retrieved using the state of art techniques like NGS, iTRAQ, qRT-PCR *etc*. Thus, the present study was formulated with the following objectives:

- Spatio-temporal and tissue specific analysis of differential expression of important genes and transcription factors associated with curcumin biosynthesis and their validation.
- Differential proteome and gene expression analysis subject to various environmental/ stress/nutrient regimes and correlation to curcumin content
- Identification of potential diagnostic markers if any for gaining insight into regulation of curcumin biosynthesis.

Chapter-1

COMPARATIVE TRANSCRIPTOME ANALYSIS AND IDENTIFICATION OF CANDIDATE STRUCTURAL GENES FOR CURCUMIN BIOSYNTHESIS

1.1. INTRODUCTION AND OBJECTIVE

Plant-derived natural compounds have been widely exploited in the development of novel drugs, with 30-50 % of contemporary medicines and nutraceuticals derived from plants (Anand et al., 2019). Secondary metabolites derived from plants have long been used as a source of biomaterial in many medical and industrial applications (Seca and Pinto, 2019; Wink, 2015). Turmeric (*Curcuma longa* L.) is a perennial, herbaceous plant, the rhizomes of which have been used as traditional medicine since 4000 years with a wide spectrum of pharmacological activities (Prasad and Aggarwal, 2011). Turmeric is abundant in tropical and sub-tropical areas and especially in Asian countries like China and India (Araújo and Leon, 2001). Indian turmeric is high in curcumin and other useful metabolites and commands a premium price in the international market, with India accounting for over 60% of global turmeric trade (Angles et al., 2011). The presence of high curcumin content in Indian turmeric makes it the best in the world (Muthusamy, 2011). The therapeutic properties of turmeric are mainly attributed to curcuminoids which is a mixture of curcumin, demethoxycurcumin, and bis-demethoxycurcumin (Figure 1). Among them, curcumin a natural polyphenolic pigment is found to be the most active one (Sharifi-Rad et al., 2020) and is popularly known as the “wonder drug of life”(Gera et al., 2017). In recent decades enormous studies were conducted to prove the biological activities of curcumin including anti-inflammatory, anti-angiogenic, antioxidant, wound healing, antimicrobial activities (Babaei et al., 2020; Wal et al., 2019). Furthermore, according to recent findings, curcumin could be used as an adjuvant drug in the treatment of COVID-19 (Pawar et al., 2021; Rattis et al., 2021).

Curcumin is a symmetric molecule also known as di-feruloyl methane (Priyadarsini, 2014a). The curcumin structure was first proposed by Polish scientists in 1910 (Miłobędzka et al., 1910) and its structure is 1,7-bis(4-hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5-dione with chemical formula $C_{21}H_{20}O_6$. Curcumin consists of two

aromatic ring system containing an o-methoxy phenolic group, connected by a seven-carbon linker consisting of an α, β-unsaturated β-diketone moiety (Grynkiewicz and Slifirski, 2012 ; Priyadarsini, 2013).

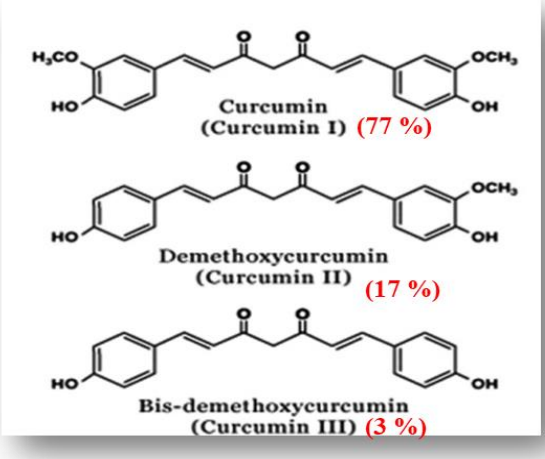


Figure 1: Different curcuminoids present in turmeric (Chempakam and Parthasarathy, 2008)

The biosynthetic pathway of curcumin gains much attention due to its pharmaceutically important properties. High curcumin content varieties are preferred by pharmaceutical industries but the curcumin content varies within turmeric accessions (Guddadarangavvanahally et al., 2002) and with agro-climatic conditions (Anandaraj et al., 2014), which is a setback for its commercial production. Curcumin is a phenylpropanoid derivative and diverse group of compounds are derived from the carbon skeleton of phenylalanine through the PPP. Putative curcumin biosynthetic pathway can be indicated in two phases: (1) upstream phase common to all the phenylpropanoids; (2) downstream phase specific to curcumin production (Figure 2).

In curcumin biosynthesis the starter substrates, cinnamoyl-CoA, p-coumaroyl-CoA, or feruloyl-CoA, are synthesized from phenylalanine by the activity of phenylalanine ammonia-lyase (PAL), 4-coumarate: CoA ligase (4CL), cinnamate-4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), cinnamate-3-hydroxylase (C3H), and O-methyltransferase (OMT). DCS catalyzes formation of diketide-CoAs by condensing p-coumaroyl-CoA and feruloyl-CoA with malonyl-CoA. CURS catalyzes formation of curcuminoids such as curcumin, demethoxycurcumin and bis-demethoxy curcumin by condensing diketide-CoAs with starter substrate.

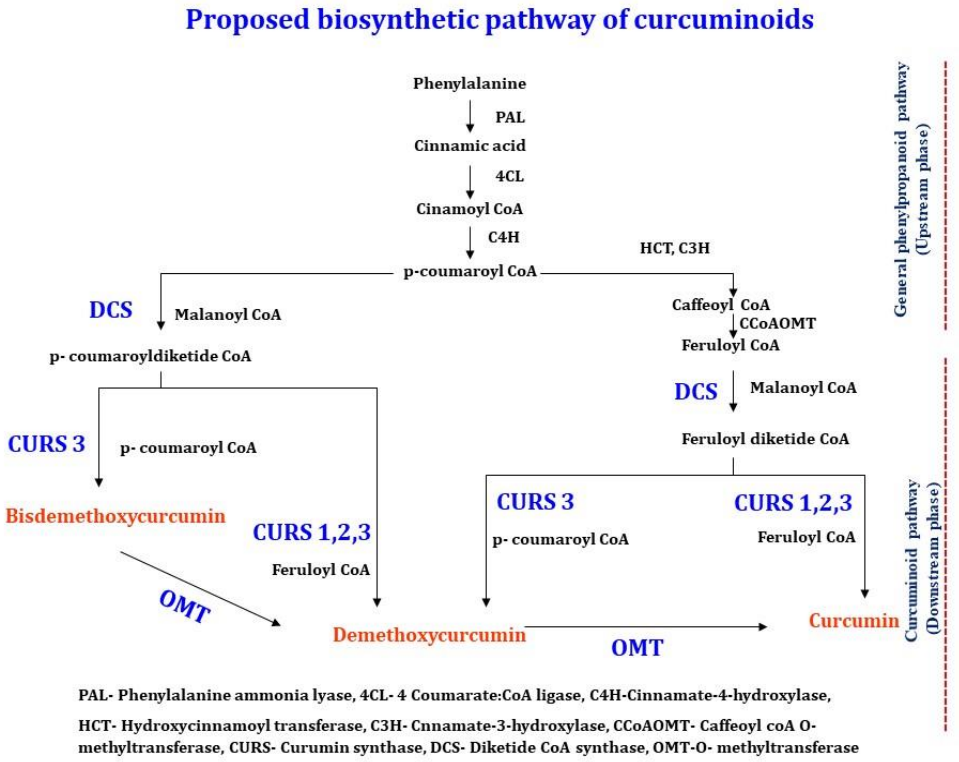


Figure 2: Proposed biosynthetic pathway of curcuminoids (Katsuyama et al., 2009)

Research on curcumin biosynthesis was initiated during 1997 by Schroder who proposed that the polyketide synthases enzymes are involved in the production of curcumin. Kita et al., 2008 suggested that curcumin is a phenylpropanoid derivative through radiotracer studies. A pathway for curcuminoid biosynthesis was proposed and novel type III polyketide synthases namely diketide-CoA synthase (*DCS*) and three curcumin synthases (*CURS1*, *CURS2*, and CURS3) were identified through genomic and biochemical approaches by Katsuyama et al., 2009 a and b. Another polyketide synthase *ClPKS10*, which is putatively involved in curcumin biosynthesis was identified by homology based RT-PCR and data mining (Resmi and Soniya, 2012). Majority of these studies involve expression analysis by designing primers based on conserved domains or random partial sequences in the public domain that may not be specific to the bonafide candidate gene involved in the pathway. Hence, they may provide an extrapolated result and need not reflect the actual expression profile of the candidate isoform. This implies the need to identify the candidate isoform(s) involved in the biosynthetic pathway since isoforms of the same gene may have different functions in different tissues and pertaining to biological stages.

Biosynthesis of curcumin requires starting precursor molecules L-phenylalanine and the first enzyme of phenylpropanoid biosynthesis, known as Phenylalanine Ammonia Lyase (PAL, EC.4.3.1.24), which catalyses the conversion of L-phenylalanine to cinnamic acid (Fig 1). Phenylalanine ammonia-lyase is the first key enzyme and the rate-limiting enzyme involved in phenylpropanoid metabolism and the entry point from primary metabolism into phenylpropanoid metabolism. (Vogt, 2010). PAL enzyme activity/PAL transcript abundance determines the flux through the PPP and the rate of phenylpropanoid production (Bate et al., 1994; Wang et al., 2014), which in turn determines the end-product levels. The upstream section of the curcumin biosynthesis pathway is common for all other phenylpropanoids and the downstream section is specific for curcuminoids (Figure 2). The structure of curcumin comprises two aromatic ring systems containing o-methoxy phenolic groups, connected by a seven-carbon linker. Curcumin and demethoxycurcumin contain three, two methyl groups respectively. According to Kita et al., 2008 O-methylation takes place after the formation of the curcuminoid skeleton (Kita et al 2008) However, it is not known when the methyl ethers at the 3' position of curcumin are incorporated, i.e., before or after curcumin scaffold formation (Katsuyama et al., 2009). There are two types of formation for curcumin either from caffeoyl-CoA to feruloyl-CoA by caffeoyl-CoA O-methyltransferase (CCOMT) which is then converted to curcumin by the action of DCS and CURS. An alternate mechanism is by direct conversion of p-Coumaroyl CoA into bis-demethoxycurcumin by polyketide synthase, which is then converted into demethoxycurcumin and then to curcumin.

Comparative transcriptome analysis is the best way to identify the candidates involved in biosynthesis pathways mainly through fold change and FPKM values of transcripts followed by functional validation using qRT-PCR (Foong et al., 2020; Jiang et al., 2020) A comparative transcriptome with samples contrasting in curcumin content (cultivated turmeric having curcumin >5.5% and a wild species *C. aromatica* (curcumin 0%) (Sheeja et al., 2015) helped in identification of a novel polyketide synthase ClPKS11 (Deepa et al., 2017; Deepa 2018) and also MYB related (Aparna et al., 2021) and bHLH/WD40 transcription factors (Mol et al., 2021). Genes so far identified in the pathway are mainly downstream genes since they are dedicated to the synthesis of curcumin, while upstream genes belong to multigene families and mostly exist as isoforms with complex structural and functional characteristics. Thus, the

curcumin biosynthetic pathway is poorly characterized with a few reports on downstream genes and little or no information on the important upstream genes. The phenylpropanoid biosynthesis route produces a wide range of compounds, owing to modifications such as acylation, methylation, hydroxylation, and gylcocylation *etc*. Furthermore, when the curcumin methyl ethers at the 3' position are incorporated, whether before or after the curcumin scaffold is produced, is uncertain.

The study was conducted with the main objective to identify differentially expressed genes involved in PPP that correlate with curcumin content.

1.2. REVIEW OF LITERATURE

1.2.1. Plant secondary metabolites and phenylpropanoids

Plant secondary metabolites are the source of medicines and The World Health Organization (WHO) informs that nearly 80% of the world's population is dependent on medicinal plants or herbs for their medical needs (Robinson and Zhang, 2011; WHO, 2019). Since ancient times dating back to 5000 years it is recorded that plants, herbs and natural compounds have been extensively used and explored for new drugs (Chandra et al., 2017). Plants are storehouses of promising elements like secondary metabolites and human population evolved over the years in close dependence on them (Robinson and Zhang, 2011). Of late an increased interest in medicinal plants is seen since 30–50% of current pharmaceuticals and nutraceuticals are plant-derived. (Anand et al., 2019). Examples of plant-derived drugs are artemisinin (Njuguna et al., 2012), vincristine (Yetgin et al., 2010) ginkgolide (Wang et al., 2006), quinine (Kremsner et al., 1994), camptothecin (Wall et al., 1966) and paclitaxel (TaxolTM) (Wani et al., 1971). Based on their biosynthetic pathway, secondary metabolites were broadly classified into terpenoids (terpenes and steroids), alkaloids, and phenylpropanoids (phenolics) (Bourgaud et al., 2001; Zhao et al., 2013). Phenylpropanoid metabolism is one of the most important metabolisms in plants, yielding more than 8,000 metabolites contributing to plant development and plant–environment interplay (Dong and Lin, 2021). Phenylpropanoids are synthesized mainly by fruits and vegetables which were reported to have diverse health benefits like anti-inflammatory, antioxidant, anticancer, and antimicrobial activities. 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 analysing the complex pathway involved in secondary metabolite biosynthesis through extensive research (Li et al., 2014)

1.2.2 Biosynthesis and regulation of phenylpropanoids

The diversity in phenylpropanoids is credited to specific enzymes and enzyme complexes that bring about regio-specific condensation, cyclization, aromatization, hydroxylation, glycosylation, acylation, prenylation, sulfation, and methylation reactions (Noel et al., 2005). Enzymes involved in the phenylpropanoid biosynthetic pathway like lyases, transferases, ligases, oxygenases were identified, many of which are encoded by gene superfamilies (Hou et al., 2013; Tohge et al., 2013; Turnbull et al., 2004; Ververidis et al., 2007; Wang et al., 2013). Three enzymes (Phenylalanine ammonia-lyase, Cinnamate 4 hydroxylases, 4 Coumaroyl CoA ligase) are required to transform L-Phenylalanine into the Coenzyme A (CoA). Activated hydroxycinnamoyl esters are capable of entering various downstream pathways as they are direct precursor for the 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 (CCoAOMT) to form Feruloyl CoA. Feruloyl CoA is identified as the precursor for Guaiacyl-lignin, Syringyl-lignin, and Coumarin biosynthesis (Baxter and Stewart, 2013; Vogt, 2010). Monocots like grasses and maize possess dual-specificity Ammonia-lyase (PAL/Tyrosine ammonia-lyase TAL) that uses tyrosine as a substrate, reducing the number of enzymes that are essential for the production of p-Coumaroyl-CoA from three in the general PPP to two (Fritz et al., 1976; Rösler et al., 1997). PAL is encoded by a multi-gene family, ranging from a few members in many species to more than a dozen copies in potato and tomato (Chang et al., 2008). Four isoforms of PAL were identified in *Arabidopsis*, five in poplar and nine in rice, each respond differentially to different stress conditions and was regulated at both developmental and spatial levels (Bhuiyan et al., 2009; Lillo et al., 2008). Most of the enzymes of the PPP are encoded by gene superfamilies, such as the cytochrome P450 membrane-bound monooxygenase (P450) gene family, the 2-oxoglutarate dependent dioxygenase (2-ODD) gene family, the NADPH-dependent reductase gene family, and the type III polyketide synthase (PKS III) gene family (Hou et al., 2013; Tohge et al., 2013; Turnbull et al., 2004; Ververidis et al., 2007; Wang et al., 2013).

Small gene families encoding enzymes of phenylpropanoids and lignin biosynthesis were identified in *Arabidopsis* and poplar (Hamberger et al., 2007). In addition to enzymatic genes, great efforts have been made to the regulation of phenylpropanoid biosynthesis. MYB, bHLH, WD40, NAC, and WRKY can co-ordinately regulate the expression of enzymatic genes at transcriptional level in a temporal and spatial manner (Deng and Lu, 2017). In maize two R2R3 MYB transcription factors (ZmMYB111 and ZmMYB148) involved in phenylpropanoid metabolism were identified. (Zhang et al., 2016). Recently NAC and MYB transcription factors involved in lignin biosynthesis were identified from *Melilotus albus* (Chen et al., 2021). Basic Helix-Loop-Helix transcription factor was identified as regulators of alkaloid biosynthesis pathway in *Coptis japonica* (Yamada et al., 2011).

Anthocyanins are the most important and extensively studied phenylpropanoids in plants and the techniques and methodologies used in deducing the anthocyanin pathway genes can be adopted for curcumin biosynthesis pathway too. All the structural genes in anthocyanin biosynthetic pathway were identified in many crops (Guo et al., 2019, 2014; Liu et al., 2019; Ma et al., 2009). MYB transcription factors (TFs) are dominant members in the regulatory network of anthocyanin biosynthesis (Yan et al., 2021). And other transcription factors such as NAC (Zhang et al., 2020), WRKY, bzip (An et al., 2017) regulates accumulation of anthocyanin in apple and WRKY transcription factors mediate regulation in grape (Amato et al., 2017). DNA-binding R2R3-MYB transcription factors, basic helix–loop–helix (bHLH) transcription factors, and WD40 repeat proteins are known to form MYB-bHLH-WD repeat (MBW) complexes, which activates the transcription of structural genes in the anthocyanin pathway in many plants including Arabidopsis (Pesch et al., 2015), carrot (Kodama et al., 2018), capsicum (Stommel et al., 2009), blueberry (Zhao et al., 2019), pear (Cui et al., 2021), Norway spruce (Nemesio-Gorriz et al., 2017). Some research has discovered that miRNAs play a function in the control of secondary metabolic pathways (Gupta et al., 2017). In addition, plant miRNAs, such as miR828, miR858 have been shown to play critical regulatory roles in phenylpropanoid biosynthesis (Sharma et al., 2016; Tirumalai et al., 2019). The accumulation of anthocyanins in the stems of *A. thaliana* is regulated by miR156-targeted squamosa promoter binding protein-like (SPL) gene and anthocyanin accumulation and activity-induced flavonols are aided by high miR156 activity. This research also shows that SPL9 inhibits anthocyanin accumulation by destabilising the MYB-bHLH-WD40 transcriptional activation complex (Gou et al.,

2011). Curcumin is a natural polyphenol synthesized through phenylpropanoid biosynthesis pathway from *C. longa* (turmeric). Due to the presence of curcumin, turmeric has been ascribed a plethora of therapeutic pharmacological qualities.

1.2.3. Techniques applied to decipher the biosynthetic pathway of secondary metabolites

1.2.3.1. Precursor feeding studies

Precursor feeding has been a successful approach for enhanced production of secondary metabolites from plant cells grown *in vitro* through overexpression of the genes involved in the biosynthetic pathway (Sivanandhan et al., 2014). This method is utilized to decipher the metabolic pathways. Biochemical studies using isotope feeding have suggested that plants synthesize salicylic acid from cinnamate due to the activity of phenylalanine ammonia-lyase (PAL) (Chen et al., 2009). Feeding experiments of *Ophiorrhiza pumila* was conducted to find which intermediate (strictosidine or strictosidinic acid) is most likely to take part in the CPT biosynthesis pathway and results showed that strictosidine, rather than strictosidinic acid, was the exclusive intermediate involved in CPT biosynthesis (Yang et al., 2021). Kita et al., 2008 developed an *in vitro* culture system of turmeric plants for feeding 13C-labeled precursors in order to explore the production of curcuminoid in the rhizomes of turmeric (*Curcuma longa* L.). This study findings point to the idea that two cinnamoyl CoAs and one malonyl CoA were used in the pathway to curcuminoids, with hydroxy and methoxy-functional groups on the aromatic rings added after the curcuminoid skeleton was formed (Kita et al., 2008).

1.2.3.2. Transcriptome and comparative transcriptome analyses

A transcriptome is a snapshot of the gene expression in a given cell or tissue at a given moment provided by capturing the total RNA within that tissue (Ward et al., 2012). The term "transcriptome" was first used by Charles Auffray (Piétu et al., 1999), refers to an organism's entire repertoire of ribonucleic acid (RNA) molecules in a specific cell type or tissue at a certain developmental stage and/or under a specific physiological condition (Dong and Chen, 2013). The term "transcriptomics" refers to the study of RNA transcription and expression levels, as well as their functions, locations, trafficking, and degradation. Important information on transcription start

sites, 5′ and 3′ sequence regions, patterns of RNA splicing and post transcriptional modifications of genes may be accrued (Wang et al., 2009). Transcriptomics covers all types of transcripts, including messenger RNAs (mRNAs), microRNAs (miRNAs), and different types of long noncoding RNAs (lncRNAs). These techniques have become increasingly practical in non-model species as the cost of high throughput sequencing has decreased due to automation and efficiency (Ward et al., 2012). Differential display reverse transcriptase-polymerase chain reaction (DDRT-PCR) was one of the first methods for the analysis of differential gene expression on the scale (Liang and Pardee, 1992) but labour-intensive and can yield results that are not reproducible (Malhotra et al., 1998). Additional technologies capable of analysing the expression of multiple genes have been developed to address the growing interest in transcriptome analysis. Two technologies developed, were cDNA microarrays (Schena et al., 1995) and serial analysis of gene expression (SAGE) (Velculescu et al., 1995). cDNA microarray data are continuous and do not directly give information about sequence variation. In SAGE, a library of joined cDNA fragments is generated and sequenced with the Sanger method (Velculescu et al., 1995). The analysis of SAGE data was limited by sequencing cost, labor, and the availability of genome sequences or ESTs. A new method cDNA AFLP (Bachem et al., 1996) was introduced and later massively parallel signature sequencing (MPSS) was developed

(Brenner et al., 2000), which produced data similar to that of SAGE, yielding very short sequencing reads (initially 32-mers) that could be aligned to a reference genome or set of ESTs. MPSS is more promising than SAGE methods due to its added advantage of increased sequencing depths. However, the library preparation is a bit complex involving bead cloning stage, which made the technology not very popular. Three dominant sequencing methods have emerged; Illumina's method of highly parallel sequencing by synthesis; SOLiD sequencing by ligation; and Roche 454's pyrosequencing that produce a tremendous volume of high-quality data. For the Illumina sequencing, the user can currently select the length of reads in the range of 36 nt to 150 nt that can be sequenced either from one end of a DNA fragment (single-end reads) or from both ends of a DNA fragment (paired-end reads). Longer reads and paired-end reads are typically selected in *de novo* assembly projects, but shorter reads are sometimes chosen for alignment to a reference genome for practical applications. In the SOLiD system, the user can currently choose read lengths of 35 nt to 75 nt in length in either single-end or paired-end format. The SOLiD system sequences two bases at a

time (thus there are 16 possible combinations to query), and any single base must be sequenced twice to identify the true sequence at a single position. However, in case of *de novo* sequencing, this 2-base encoding system is a drawback. Roche's 454 sequencing provide wider sequence length distribution, and sequences are encoded in normal nucleotide space. Most 454 reads are now longer than 500 nucleotides, with a mode around 700 nucleotide and a maximum length over 1000 nt. 454 sequencers yield high-quality transcriptome assemblies, but these data sets have much lower depth and is costlier.

Transcriptome sequencing, also known as RNA-Seq, is a recently developed technique that quantifies transcriptome in a given sample and thus promotes analysis of gene expression (Wang et al., 2009). Total or fractionated, RNA such as poly(A)+) is converted to a library of cDNA fragments with adaptors attached to one or both ends and each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing), giving a read length of 30-400 bp. In principle, any high-throughput sequencing technology (Holt and Jones, 2008) can be used for RNA-Seq, and the Illumina IG (Lister et al., 2008; Nagalakshmi et al., 2008; Wilhelm et al., 2008), Applied Biosystems SOLiD (Cloonan et al., 2008) and Roche 454 Life Science (Barbazuk et al., 2007; Vera et al., 2008) systems have already been applied for this purpose. Recent RNA-Seq studies for the exploration of biosynthetic pathways in spices include those in black pepper (Hao et al., 2016; Hu et al., 2015; Jin et al., 2018), ginger (Jiang et al., 2017) turmeric (Annadurai et al., 2013; Deepa et al., 2017; Sheeja et al., 2015) and chilli pepper (Gao et al., 2021; Jarret et al., 2019; Kang et al., 2020; Liu et al., 2013; Martínez et al., 2021)

Comparative transcriptome is the best way to identify candidate genes involved in biosynthesis pathways. Several comparative transcriptomes were developed to mine the candidate genes of the biosynthetic pathways in many crops (Foong et al., 2020; He et al., 2018; Liu et al., 2015; Soltani Howyzeh et al., 2018). In spices, several studies have been conducted to explore pathway genes (Deepa et al., 2017; Hao et al., 2016; Khew et al., 2020; Liu et al., 2013) as secondary metabolites are the most important molecules that impart the nutritional and therapeutic properties to spices. Curcumin content was shown to be present in significantly different concentrations in PC and PS. When we subjected plant to nutrient stress, initially the number of secondary metabolites increased and later decreased due to stress. When black glutinous rice

seedlings were exposed to salt stress for 2-4 days, the highest levels of anthocyanin were found, while stress for up to 6-8 days inhibited anthocyanin accumulation. (Chunthaburee et al., 2016).

The transcriptome sequencing outputs and functonal annotation results obtained (total of unigenes obtained 138,015 unigenes, with 46.94% of mean GC of the transcript) are comparable to the recent studies on *Trachyspermum ammi* (L.) (123,488 unigenes; 38.35% of mean GC of the transcript) (Soltani Howyzeh et al., 2018). In *Curcuma longa* transcriptome datasets 100% of the unigenes were annotated using NCBI non-redundant protein database using BLASTX. In *Pueraria thomsonii* Benth, among the 44,339 nonredundant transcripts obtained, 43,195 could be annotated (He et al., 2019). Zhang et al., 2018 reported that among 293,823 transcripts only 190,567 unigenes could be annotated in Jerusalem artichoke. In the *I. balsamina* transcriptome dataset, 26.04% of the unigenes did not possess significant similarity to sequences from other crops (Foong et al., 2020), which is close to the percentage of 'orphans' or 'taxonomically restricted genes' (TRGs) in a given species (Wilson et al., 2005). To confirm the robustness of the *Curcuma longa* transcriptomes, qRT-PCR of 11 selected genes and 8 TFs were performed and the analysed results indicated reliability of the transcriptome data. Several biosynthetic pathway candidate genes were identified from transcriptome and validated using qRT-PCR technique. (Foong et al., 2020; Gao et al., 2020; Lin et al., 2019; Zhao et al., 2021).

Biosynthesis pathway of curcumin is divided into two phases; upstream phase common to all phenylpropanoid compounds and downstream phase specific to curcumin biosynthesis. All the genes of the pathway were sucessfully identified by differential transcriptome analysis based on the FPKM values. DEG analysis revealed that majority of the genes involved in the curcumin biosynthesis pathway were expressed at significant levels under control condition compared to stressed condition (Deepa et al., 2017).

1.2.3.3. Proteomics analysis

Proteomics is the study of the identification and quantification of overall protein content in a cell, tissue, or organism using various technologies. It works in tandem with other "omics" technologies like genomics and transcriptomics to determine the

identity of an organism's proteins and to understand their structure and functions (Aslam et al., 2017). Protein structure and functions were characterised using both analytical and bioinformatics approaches in proteomic study. MALDI-TOF-MS and 2-D gel electrophoresis were employed as analytical techniques. There are different advanced methods available in proteomics like Isotop-coded affinity tags (ICAT), Isobaric Tags for Relative and Absolute Quantification (iTRAQ), Absolute quantification (AQUA), ESI-Q-IT-MS, SELDI-TOF-MS. A comparative proteome analysis in rice to map the starch biosynthetic pathway genes by 2D PAGE/MS to obtain a comprehensive overview of starch biosynthesis and identification of important protein markers involved in starch biosynthesis (Chang et al., 2017). The results of a combined proteomics and metabolomics study in *Arabidopsis* in response to glucosinolate biosynthesis perturbation revealed that metabolomics data correlated with that of proteomics and revealed intimate molecular connections between cellular pathways/processes and glucosinolate metabolism (Chen et al., 2012). An iTRAQ-based quantitative proteomics-based analysis of black rice grain reveals metabolic pathways linked to anthocyanin biosynthesis, as well as proteins involved in ACN biosynthesis and sugar synthesis. The gene expression was upregulated, particularly from the onset of black rice grain development. Genes and proteins related to signal transduction, redox homeostasis, photo synthesis and N-metabolism decreased during grain maturation. In another study differentially expressed proteins that are in response to light with the capitulum development of the chrysanthemum 'Purple Reagan' were identified using iTRAQ technique, and the studies could establish a correlation between the proteomic and the transcriptomic data. A novel GDSL esterase APG was identified to play an important role in light signal transduction and a putative mechanism of light-induced anthocyanin biosynthesis in *Chrysanthemum* was proposed (Hong et al., 2019).

1.2.3.4. Metabolomics approach

Metabolomics, is a novel strategy for characterization of metabolite profiles and is being recognised as a powerful tool for discovery and production of secondary metabolites (Breitling et al., 2013). Recent technological breakthroughs in precision mass spectrometry and data interpretation have revolutionized metabolomics experiments. The simultaneous and rapid screening of hundreds of metabolites from a wide range of chemical classes is possible with activity-based and global metabolomic profiling methodologies, making them ideal tools for uncovering new enzyme activity

and metabolic pathways (Prosser et al., 2014). Li et al., 2016 reported the use of metabolomics to study the metabolic pathways of the etiolation of fresh-cut Chinese water chestnuts. Metabolomic approaches were conducted to identify differences in metabolite profiles between transformed T8; tartary buckwheat and T10; common buckwheat hairy roots in *Fagopyrum tataricum Gaertn* (Thwe et al., 2013). Differential metabolome analysis was also utilized to investigate metabolite production in various physiological states. To explore the developmental and environmental regulation of secondary metabolism in *Smallanthus sonchifolius* (yacón), a comparative metabolic analysis was conducted in different developmental stages of the plant (Padilla-González et al., 2019). Another study in *Taxus* species, a pharmacologically significant crop, reveals differences in taxoids and flavonoids among three *Taxus* species based on comparative metabolomic analysis (Zhou et al., 2019). Several investigations have been undertaken using a combined transcriptome and proteome analysis. (Lu et al., 2021; Pan et al., 2021; Zhang et al., 2021).

1.2.4. Turmeric and commercial importance

Turmeric (*Curcuma longa* L.) is an herbaceous rhizomatous perennial plant that generally attains a height of 3-5 feet with green-coloured oblong leaves, yellowish funnel-shaped flowers, and pseudo stem at the base. The pseudo stem is modified into mother rhizome, which branches into primary, secondary and tertiary rhizome (Sasikumar, 2005a) (Figure 3). It is grown as an annual, on raised flat beds or ridges for ease in harvesting of rhizomes (Hussain et al., 2015). Turmeric is originated in the Indo-Malayan region (Purseglove, 1968) and widely distributed in the tropics of Asia (Gupta et al., 2013). Turmeric can be grown up to 1,600 meters above sea level (Lal, 2012) and can grow in diverse environmental situations at a temperature of 20–30°C (Prasad and Aggarwal, 2011) with an yearly rainfall of 1500 mm (Rahi et al., 2020). It grows in well-drained sandy or clay loam soils, having a pH of 4.5–7.5 with good organic status, where it flourishes outstandingly (Nair, 2013). The rhizome is the commonly used part that measures 2.5-7.0 cm (in length), and 2.5 cm (in diameter) with segmented skin (Lal, 2012). Turmeric is a cross pollinated triploid ($2n = 3x = 63$) species that is vegetatively propagated *via* underground rhizomes (Sasikumar 2005b). A major impediment to the crop improvement programs in turmeric is the low seed set due to self-compatibility. So clonal selection, polyploidy induction, mutation breeding, and molecular assisted breeding methods are generally used for the genetic

improvement of turmeric (Ravindran et al., 2007). A viable seed set can be obtained under certain conditions, which enables recombination breeding through hybridization and open-pollinated progeny selection (Sasikumar, 2005). IISR Prabha and IISR Prathibha, were the first open-pollinated progenies of turmeric (Sasikumar et al., 1996).

Figure 3: (A) Field view of *Curcuma longa* (turmeric), (B) Whole plant, (C) leaf, (D) pseudo-stem, (E) flower, (F) primary rhizome, (G) root, (H) turmeric clump, (I) turmeric powder

Turmeric is also an important commercial crop (Babu, 2015)and major cultivators of turmeric are Andhra Pradesh, Maharashtra, Orissa, Tamil Nadu, West Bengal, Karnataka, and Kerala. In India, the climatic conditions, fertility of the soil, rainfall, and cultivation practices are highly favourable for the cultivation of turmeric. Countries like USA, UK, and Japan exhibit an added interest in Indian turmeric due to the high value given to it by quality consciousness Indian exporters. India has a major

share in the production, consumption, and export of turmeric in the world. The oldest reference of turmeric usages was in the 'Atharvaveda' a holy book of Hindus as a remedy to cure skin diseases, jaundice, and greying of hair (Ravindran et al., 2007). Turmeric is used in the cosmetic industry (Arct et al., 2014; Gonçalves et al., 2014; Wongthongdee and Inprakhon, 2013) and may be the first known cosmetic as it has been traditionally smeared on the skin by women (Gopinath and Karthikeyan, 2018), as a food preservative (Arulkumar et al., 2017) and in traditional systems of medicine (Ayurveda, Siddha and Unani, Tibetan) (Ravindran et al., 2007).

1.2.4.1. Turmeric and Curcuminoids

Curcumin was first isolated in 1815 by Vogel and Pelletier, and purified crystalline compound was obtained in 1870 by Daube (Daube, 1870) and the first paper on the synthesis of curcumin was reported by Lampe in 1918. The chemical formula is $C_{21}H_{20}O_6$ and the chemical structure is diferuloylmethane or 1, 6-heptadiene-3,5-dione-1,7-bis (4-hydroxy-3- methoxyphenyl) was first proposed by Polish scientists in 1910 (Milobędzka et al., 1910). Curcumin (chemical formula $C_{21}H_{20}O_6$) is a diferuloylmethane having a shiny yellow-orange color, with a molecular weight of 368.39 g/mol and melting temperature of 183°C. Curcumin can be found in two different tautomeric forms: keto and enol and in both solid state and in solution, the enol form is more energetically stable (Akram et al., 2010). The structure of curcumin comprises two aromatic rings containing o-methoxy phenolic groups, connected by a seven-carbon linker consisting of an α, β-unsaturated β-diketone moiety. The structure of curcumin is given in Fig.2. Curcumin has three reactive functional groups: one diketone moiety, and two phenolic groups (Priyadarsini, 2013). It is almost insoluble in water with a logP value of ~3.0 and soluble in polar solvents like DMSO, methanol, ethanol, acetonitrile, chloroform, ethyl acetate, *etc*. It is sparingly soluble in hydrocarbon solvents like cyclohexane and hexane (Priyadarsini, 2014b).

Curcumin is a wonder drug and is used as a preventive measure for various health conditions including cardiovascular (Li et al., 2020) and diabetic problems, Alzheimer's, cancer, osteoarthritis, and pancreatitis (Gregory et al., 2008; Henrotin et al., 2010; Lin et al., 1997, 1998; Seo et al., 2011). Even though, the medical efficacy of the native curcumin is poor, because of low aqueous solubility, chemical instability, photodegradation, rapid metabolism, and short half-life (Her et al., 2018). As a result, a

significant number of bioavailable curcumin-based formulations were introduced with a varying range of enhanced bioavailability. Several approaches such as inhibition of curcumin metabolism with adjuvants like piperine that interferes with glucuronidation as well as novel solid and liquid oral delivery systems, curcumin nanoparticles, use of curcumin phospholipid complex, and use of curcumin structural analogues have been attempted to improve absorption of curcumin, bioavailability and rapid expulsion from the body (Anand et al., 2007; Cas and Ghidoni, 2019).

Curcumin has a wide range of therapeutic actions. Over 18,000 published regular manuscripts, reviews and meta-analyses revealed the potential health benefits of curcumin with half of them appearing after 2014 (Yeung et al., 2019). Most of the reported works were targeted towards the potential role of curcumin against different chronic diseases and conditions associated with inflammation and oxidative stress (Yeung et al., 2019). Compared to the risk involved in the prevention and treatment of chronic diseases curcumin have lots of advantages. Curcumin has a wide range of therapeutic actions such as hepatotoxicity, cardiotoxicity, nephrotoxicity, pulmonary fibrosis, inflammatory bowel disease, ulcers, neoplastic conditions, and multiple drug resistance (Yavarpour-Bali et al., 2019).

1.2.4.2. Genotype x Environment interaction on curcumin status

The content of curcuminoids in turmeric rhizomes vary often with varieties, locations, sources, and cultivation conditions (Li, 2011). Curcumin content has been reported to vary between *C. longa* accessions (Wang et al., 1999; Jayaprakasha et al., 2002; Sandeep et al., 2016). Tamil Nadu Agricultural University, Coimbatore conducted a study with fifteen turmeric genotypes (Shanmugasundaram et al., 2001). Curcumin levels varied from 1.23 to 5.67%, with RH-5 having the greatest value. *C. longa* rhizomes collected from ten different locations of Thailand showed differential accumulation of curcuminoids (Pothitirat and Gritsanapan, 2005). The fact the differential performance of genotypes across environments is the major challenge in spice industry, which demands for consistent curcumin and yield. Previous studies have mentioned that curcumin content varies from place to place due to genetic composition and influence of environment and agroclimatic conditions (Rama Rao and Rao 1994; Singh et al., 2013). A study was undertaken in 2014 to examine varietal stability *vis a vis* curcumin production and yield. Eleven cultivars were tested for fresh yield, curing

percent, curcumin, and dry yield (five environments) in ten different environments across India. Environments were responsible for a considerable share (70.8%) of variation in fresh yield, curing percent (31.2%), curcumin content (17.7%), and dry yield (15.7%). Results on curcumin and curing percent showed that IISR Kedaram was a consistent variety accumulating more or less similar curcumin levels across five different environments and is considered as the most stable cultivar for curcumin production. Results on curcumin production showed that IISR Kedaram was the most consistent cultivar for curcumin production across five environments (Anandaraj et al., 2014). A prediction model based on artificial neural networks (ANN) was developed in which 11 parameters of soil and meteorological factors were provided as input to obtain curcumin content as output. This ANN model has been demonstrated may be used for identifying the ideal site and area for maximum curcumin (Akbar et al., 2016). Another study evaluated the impact of climatic conditions on the curcumin content and the color of turmeric. IISR-Alleppey finger turmeric rhizomes were planted in the five selected villages of Central Kerala with varying climatic conditions. Results showed that the soil parameters have a minor effect on the color and curcumin content of turmeric. Furthermore, higher elevations in Kerala, where the average temperature is between 20 and 25 degrees Celsius, are perfect for growing turmeric with a high color value (Peter 2020). Another study was conducted with samples of turmeric rhizomes collected from various districts of Maharashtra and its curcumin content was determined. According to the studies mentioned above, environment has a significant impact on the curcumin content. Curcumin content of rhizomes increased due to Boron deficiency, while the overall rhizome yield and curcumin yield decreased (Dixit et al., 2002).

1.2.5. Important genes involved in Curcumin biosynthesis pathway

1.2.5.1. Phenylalanine ammonia lyase (PAL)

Aromatic amino acid deaminases are vital enzymes in plants that enable the transfer of carbon from primary to secondary metabolism (Barros and Dixon, 2020). Phenylalanine ammonia-lyase (PAL; EC 4.3.1.5) is an aromatic amino acid deaminase that connects the primary metabolism into the phenylpropanoid metabolism in plants (Hahlbrock and Scheel, 1989). PAL catalyzes the nonoxidative deamination of L-phenylalanine to trans-cinnamic acid and ammonium ion (Figure 4), the first committed step in the PPP. PAL belongs to a large amino acid ammonia-lyase/aminomutase super-

family, wherein it is phylogenetically clustered to the histidine ammonia-lyases (HALs) and tyrosine ammonia-lyases (TALs) (Nugroho et al., 2002). PAL plays a critical function in integrating plant primary metabolism and phenylpropanoid metabolism, hence research on PAL has traditionally acquired a lot of interest (Janas, 1993).PAL from dicotyledonous plants predominantly and efficiently deaminates L-phenylalanine, while PAL from yeast and some monocots, e.g., maize, converts L-phenylalanine and L-tyrosine to their corresponding cinnamic acid and p-coumaric acid (Fritz et al., 1976 ; (Rösler et al., 1997). PAL was identified several decades ago in higher plants and microorganisms and some fungi and a few bacteria, but not in animals (Fritz et al., 1976; Bevan et al., 1989; Estabrook and Sengupta-Gopalan, 1991; Minami et al., 1989).

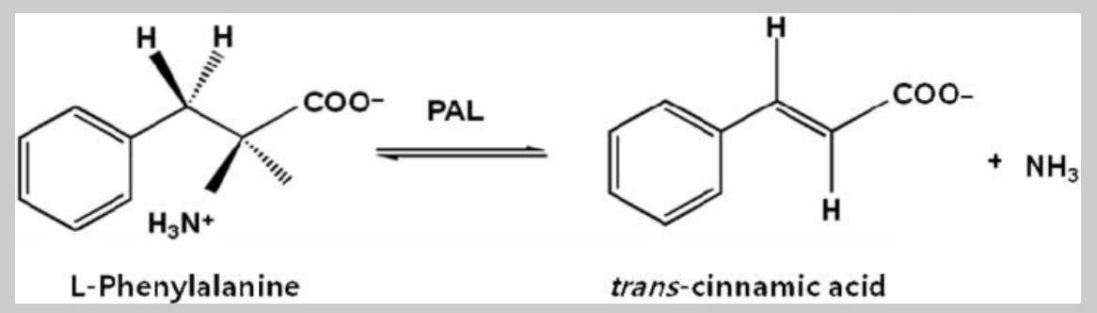


Figure 4: Deamination of L-Phenylalanine to trans-cinnamic acid by Phenylalanine ammonia lyase (PAL)

Deamination of phenylalanine is generally considered to represent a key point at which carbon flux into this pathway is controlled (Kumar and Ellis, 2001). The broad variety and massive quantities of phenylpropanoid compounds found in plant materials highlight the relevance of this enzyme in plant metabolism (Jones, 1984). PAL was first discovered in 1961 by Koukol and Conn (1961), (Kosuge and Conn, 1962) and has been studied extensively in many plants such as rice (Minami et al., 1989), Arabidopsis (Cochrane et al., 2004; Wanner et al., 1995), loblolly pine (Whetten and Sederoff, 1992), *Salvia miltiorrhiza* (Song and Wang, 2009), *Lycoris radiata* (Jiang et al., 2011), *Ginkgo biloba* (Xu et al., 2008), *Juglans regia* (Xu et al., 2012), Eucalyptus (Carocha et al., 2015), watermelon (Dong and Shang, 2013) and *Melissa officinalis* (Weitzel and Petersen, 2010). In most of the plants studied so far, PAL is encoded by a multi-gene family, presented as multiple isoforms (Cramer et al., 1989; Wanner et al., 1995). The relevance of this diversity is unknown, however evidence of metabolic channelling within phenylpropanoid metabolism implies that partitioning of photosynthates into different branches of phenylpropanoid metabolism could require labile multienzyme

complexes including unique PAL isoforms (Hrazdina and Wagner, 1985). Bolwell et al., 1985 was the first one to demonstrate multiple PAL isoforms in *Phaseolus vulgaris*. (Jahnen and Hahlbrock, 1988) found that the de novo synthesis of PAL isoforms is induced by biotic and abiotic factors and so, the differential expression of PAL isoforms leads to the synthesis and accumulation of phenolic compounds in different tissues of phenylpropanoid products. In *Arabidopsis thaliana*, four genes have been annotated as provisionally encoding PAL and all four putative *Arabidopsis PAL* genes were cloned, with the corresponding recombinant proteins expressed, purified to apparent homogeneity, and characterized concerning kinetic and physical parameters (Cochrane et al., 2004). In *P. crispum*, PAL is encoded by a family of four genes and three of them appear to be activated by fungal elicitors (Lois et al., 1989). A crystal structure became available for a codon-optimized, heterologously expressed enzyme from *Petroselinum crispum*. The *PAL* gene family in poplar (*Populus trichocarpa*) consists of five genes (PtrPAL1–5), which are separated into two groups by phylogenetic analysis (Tsai et al., 2006). In *Solanum tuberosum*, the genomic complexity and structural comparability of the *PAL* gene were investigated. It has about 40-50 *PAL* genes per haploid genome, and high cDNA heterogeneity suggests that at least ten, if not more, of these genes, are active (Joos and Hahlbrock, 1992). In potatoes, more than 40 copies were reported (Joos and Halbrook, 1992). PAL appears to exist as a gene family in all higher plants, and the presence of PAL isoforms is a widespread occurrence (Rasmussen and Dixon, 1999). *PAL* genes have been less extensively studied in gymnosperms than in angiosperms and in *Pinus taeda*, a single *PAL* gene was identified (Bagal et al., 2012). Angiosperm PAL is encoded by multigene families in which some members are differentially expressed throughout the plant and in response to environmental stimuli (Liang et al., 1989; Logemann et al., 1995).

PAL is a critical enzyme in plant stress response, in addition to its important role in plant development. Pathogenic attack, tissue injury, UV irradiation, low temperature, or low nitrogen, phosphate, or iron levels all enhance its production (Dixon and Paiva, 1995). Distinct *PAL* genes are assumed to be involved in different branches of the PPP, and the encoded proteins form homo- or hetero-tetramers. (Cochrane et al., 2004; Reichert et al., 2009; Tsai et al., 2006). *In vivo*, PAL occurs as a tetrameric enzyme with a molecular weight of 275–330 kDa (Zimmermann and Hahlbrock, 1975; Appert et al., 1994). From the suspension cell cultures of bean (*Phaseolus vulgaris* L.), several

tetrameric forms with some varying molecular weights, isoelectric point values, and substrate binding affinities were purified, suggesting that PAL can form heterotetramers (Bolwell et al., 1985). The recombinant proteins can form heterotetramers when distinct tobacco PAL isogenes are co-expressed in *E. coli*. (Reichert et al., 2009).

Phenylalanine ammonia-lyase in ginger and turmeric

The activities of phenylalanine ammonia-lyase in crude protein extract derived from young growing leaves, shoots, and rhizomes of ginger and turmeric were investigated. This experiment used [U-14C] L-Phe as substrate and monitored the formation of [U-14C] cinnamic acid as a product. All crude protein extracts had phenylalanine ammonia-lyase enzyme activity. The results of these assays demonstrated that PAL activity was generally higher in ginger tissues than in turmeric. Moreover, the highest activity in ginger was found in crude extracts from developing leaves, which was significantly higher than PAL activity in developing shoots. In contrast, the highest activity in turmeric was in the developing rhizomes, which possessed PAL activity that was about 9-fold higher than in developing leaves or shoots (Ramirez-Ahumada et al., 2006). Following *Pythium myriotylum* infection, researchers looked for a link between phenylalanine ammonia-lyase and tyrosine ammonia-lyase activities and phenolics and curcuminoid concentration in ginger and its wild relative *Zingiber zerumbet*. Even though the intrinsic PAL specific activity of the ginger cultivar was higher than that of the wild taxon, after infection of rhizomes with *Pythium myriotylum*, a steady drop in both PAL and TAL activities was observed. (Ganapathy et al., 2016).

Structural Properties of PAL Proteins

The first three-dimensional structure of phenylalanine ammonia-lyase (PAL) from *Rhodosporidium toruloides* has been identified at 2.1 resolution (Calabrese et al., 2004). PAL is a predominantly a-helical protein with 52% of the residues in 23 a-helices and only 5% of the residues in eight b-strands (Ritter and Schulz, 2004). PAL is a helix-containing protein having two extra structural segments: a mobile N-terminal extension, which possibly anchors the enzyme and interacts with other cellular components, and a specific shielding domain sited over the active centre, which probably controls enzyme activity by restricting the access of the substrate to a narrow channel so that the risk of inactivation by nucleophiles in conjunction with dioxygen is minimized (Calabrese et al., 2004; Ritter and Schulz, 2004) (Figure 5). For active catalysis, PAL and other members

of the ammonia-lyase/aminomutase superfamily uses the same co-factor, 4-methyldiene-imidazol-5-one (MIO); it is formed by the autocatalytic cyclization and dehydration of an internal tripeptide segment, namely, Ala-Ser-Gly (Alunni et al., 2003; Röther et al., 2002). The enzyme resembles histidine ammonia-lyase, presumably, phenylalanine ammonia-lyase is an offspring of HAL when fungi and plants diverged from the other kingdoms (Ritter and Schulz, 2004).

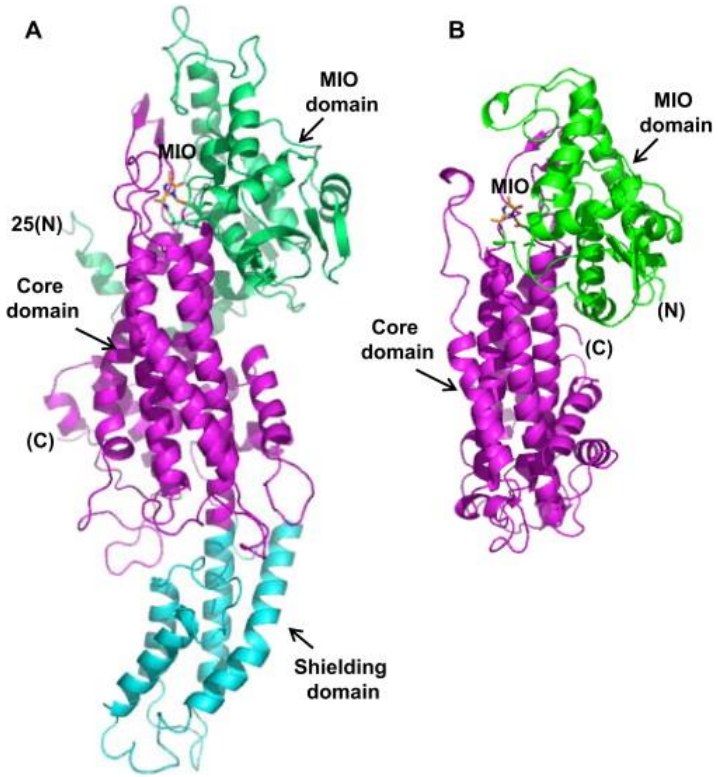


Figure 5: Comparison of the tertiary Structures of PAL Monomer from Parsley (*Petroselinum crispum*) **(A)** and HAL from *Pseudomonas putida* **(B)** (Zhang and Liu, 2015)

PAL Active Site and Enzyme Mechanism

The active site of PAL was comparable to that of histidine ammonia lyase (HAL), whose structure was determined using X-ray crystallography. (Ritter and Schulz, 2004). Hanson & Havir (Hanson and Havir, 1979, 1970) hypothesised a mechanism for the PAL reaction that began with the addition of the substrate's -αamino group to the prosthetic electrophile, which was thought to be dehydroalanine at the time. However, this method does not explain how the enzymic base can extract the nonacidic β-H of the substrate (Hermes et al., 1985) Another mechanism, including an electrophilic attack on the phenyl ring by the prosthetic group, was proposed about few years ago, and the new mechanism was supported by the ease with which alternative substrates reacted

especially 3-hydroxyphenylalanine (m-tyrosine), which facilitated the electrophilic attack (Schuster and Rétey, 1995) (Figure 6).

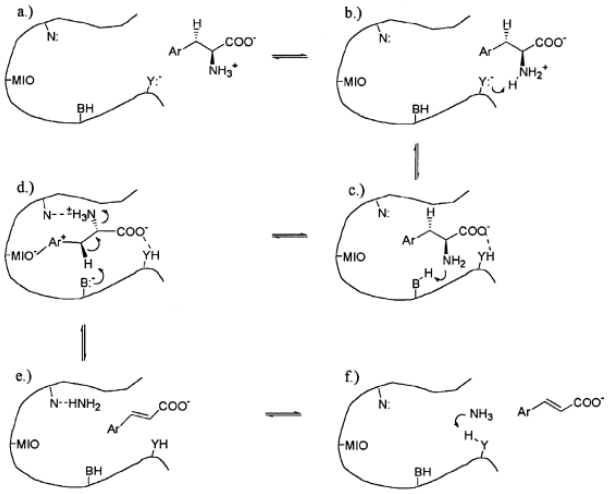


Figure 6: Proposed mode of entering the substrate (A, B and C), reaction (D) and the release of the products (E, F) for the PAL and HAL reactions. Y denotes Y110 and Y53, N denotes N260 and N195, B denotes Y251 and E414 for PAL and HAL, respectively (Rother et al.,2002).

Regulation of Phenylalanine ammonia lyase (PAL)

The biosynthesis of phenylpropanoids involves a series of central enzyme-controlled processes (referred to as the general PPP) from which branch paths lead to various aromatic end products. Genes encoding downstream enzymes of phenylpropanoid biosynthesis pathway are often co-ordinately regulated by PAL. PAL mRNA and enzyme levels are highly regulated spatially and temporally leading to a tissue-specific accumulation (Bate et al., 1994). MYB, LIM, and NTS transcription factors regulate PAL during development in different environmental conditions (Zhao and Dixon, 2011; Zhong and Ye, 2007). Moreover, post transcriptional regulation was also observed. In French bean phosphorylation of recombinant poplar PtrPAL by endogenous PAL-kinase was demonstrated to reduce PtrPAL stability (Allwood et al., 1999). Moreover, it was recently shown that in Arabidopsis, PAL activity is regulated post-transcriptionally through ubiquitination by Kelch repeat F-box (KFB) proteins to target PAL for degradation (Zhang et al., 2013) (Figure 7).

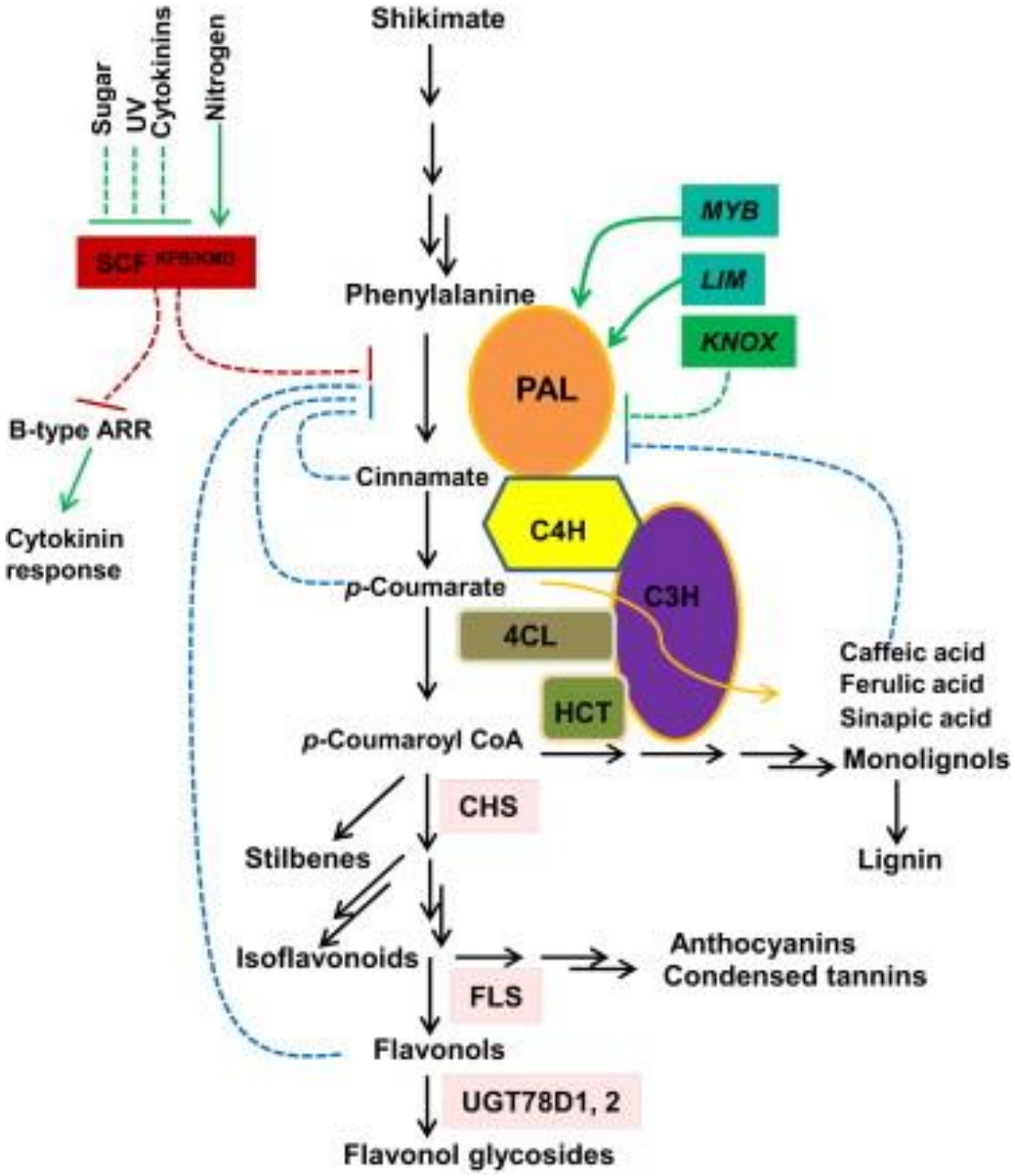


Figure 7 : Schematic diagram of the Phenylpropanoid Biosynthesis Pathway, Highlighting Molecular Regulation on the Gateway Enzyme PAL and the Physical Interaction of the Early PPP Enzymes (Zhang and Liu, 2015)

Product inhibition, transcriptional and translational regulation, post-translational inactivation and proteolysis, enzyme organization/subcellular compartmentation, and metabolite feedback regulation are all known ways for controlling PAL (Zhang and Liu, 2015). PAL has product inhibition properties ; the product of its catalysed reaction, t-CA, inhibits its activity (Appert et al., 1994; O'Neal and Keller, 1970). Transcriptional regulation of PAL also prominent. A few R2R3 MYB transcription factors have been

shown to trans-activate PAL promoters, allowing tissue-specific expression of PAL to be controlled(Martin and Paz-Ares, 1997). When AmMYB305, which is found in the carpels of early snapdragon flowers (*Antirrhinum majus*), is co-expressed in tobacco protoplasts, it activates the PAL promoter (Sablowski et al., 1994). In tobacco, a specific transcription factor, NtMYBAS1/2, binds to MYB to activate gene expression leading to phenylpropanoid synthesis in the anthers. Ubiquitination, proteolysis and phosphorylation are the major post translational modification of PAL. In eukaryotic systems, the ubiquitin-26S proteasome system dominates selective protein degradation (Vierstra, 2003). The (poly)ubiquitination of four *Arabidopsis* PAL isoforms was confirmed by immunoblotting of transiently expressed AtPAL proteins against the anti-ubiquitin antibody (Zhang et al., 2013). In eukaryotic cellular processes, phosphorylation is another main kind of post-translational modification of proteins. Bolwell (1992) reported the detection of PAL phosphorylation in suspension cultures of French bean cells (Bolwell, 1992). In assaying eight protein kinases from Arabidopsis, a calcium-dependent protein kinase, AtCDPK, could phosphorylate the short polypeptide of PAL and a poplar recombinant PAL enzyme (Cheng et al., 2001) PAL and phenylpropanoid biosynthesis activity appear to be metabolically regulated by specific biosynthetic intermediates or chemical signals, in addition to transcriptional regulation in response to environmental stimuli and nutritional balance. t-CA has long been proposed as a signal molecule, regulating flux into the pathway (Bolwell et al., 1986; Lamb et al., 1979). *PAL* gene expression and enzyme activity are negatively regulated by t-CA, either *via* exogenous application of the compound or by blocking the downstream C4H activity for conversion to cinnamic acid (Bolwell et al., 1986; Lamb et al., 1979).

1.2.5.2 O-methyltransferase

A plethora of substitution events performed by substrate- and position-specific enzymes provide diverse secondary metabolites. Plant O-methyltransferases (OMTs) are a diverse family of enzymes that methylate the oxygen atom of a variety of secondary metabolites including phenylpropanoids, flavonoids, and alkaloids. Physiological properties and chemical reactivity of secondary metabolites can change through the methylation of phenolic groups by S-adenosyl-L-methionine (SAM)-dependent O-methyltransferases (OMTs; EC 2.1.1) (Bao Ting Zhu et al., 1994). Plant S-adenosyl-L-methionine-dependent methyltransferases (SAM-Mtases) are the

important enzymes in phenylpropanoid, flavonoid, and many other metabolic pathways. (Joshi and Chiang, 1998). Two different O-methyltransferases are involved; namely the caffeic acid OMT (COMT) subfamily and the caffeoyl-CoA OMT(CCoAOMT) subfamily (Joshi and Chiang, 1998) are present with difference in the molecular weight and cation dependency of the proteins. Members of the COMT subfamily have a molecular weight ranging from 40-43 kDa and cation independent, while members of CCoAOMT subfamily are smaller in size (26–30 kDa) and are cation-dependent (Liu et al., 2020b).

Caffeic acid OMT (COMT)

Caffeic acid O-methyltransferase (COMT; EC 2.1.1.68) catalyses predominantly the O-methylation of the 5-hydroxyl group of 3-methoxy-4,5-dihydroxy precursors, preferentially as the aldehyde or alcohol (Figure 8).

The COMTs are plant type-1 family of SAM-dependent O-methyltransferases (Noel et al., 2003) having ~360 amino acid residues with an auxiliary N-terminal domain functioning in homodimerization, and are independent of metal ions for carrying out their function (Louie et al., 2010). One or more representatives of COMT occur in almost all plant lineages. COMTs from alfalfa (*Medicago sativa*), wheat (*Triticum aestivum*), and aspen (*Populus tremuloides*) (Li et al., 2000; Ma and Xu, 2008) have great affinity to 5-hydroxyconiferaldehyde and caffeoyl aldehyde, and lesser affinity to substrates like 5-hydroxyconiferyl alcohol and caffeoyl alcohol, while substrates like 5-hydroxyferulic acid and caffeic acid are the poorest indicating a greater affinity to aldehydes followed by alcohols and acids. In addition, COMTs promote O-methylation at 5-hydroxyl group than 3-hydroxyl group with a complete lack of activity at the 4-hydroxyl group (Bhuiya and Liu, 2010).

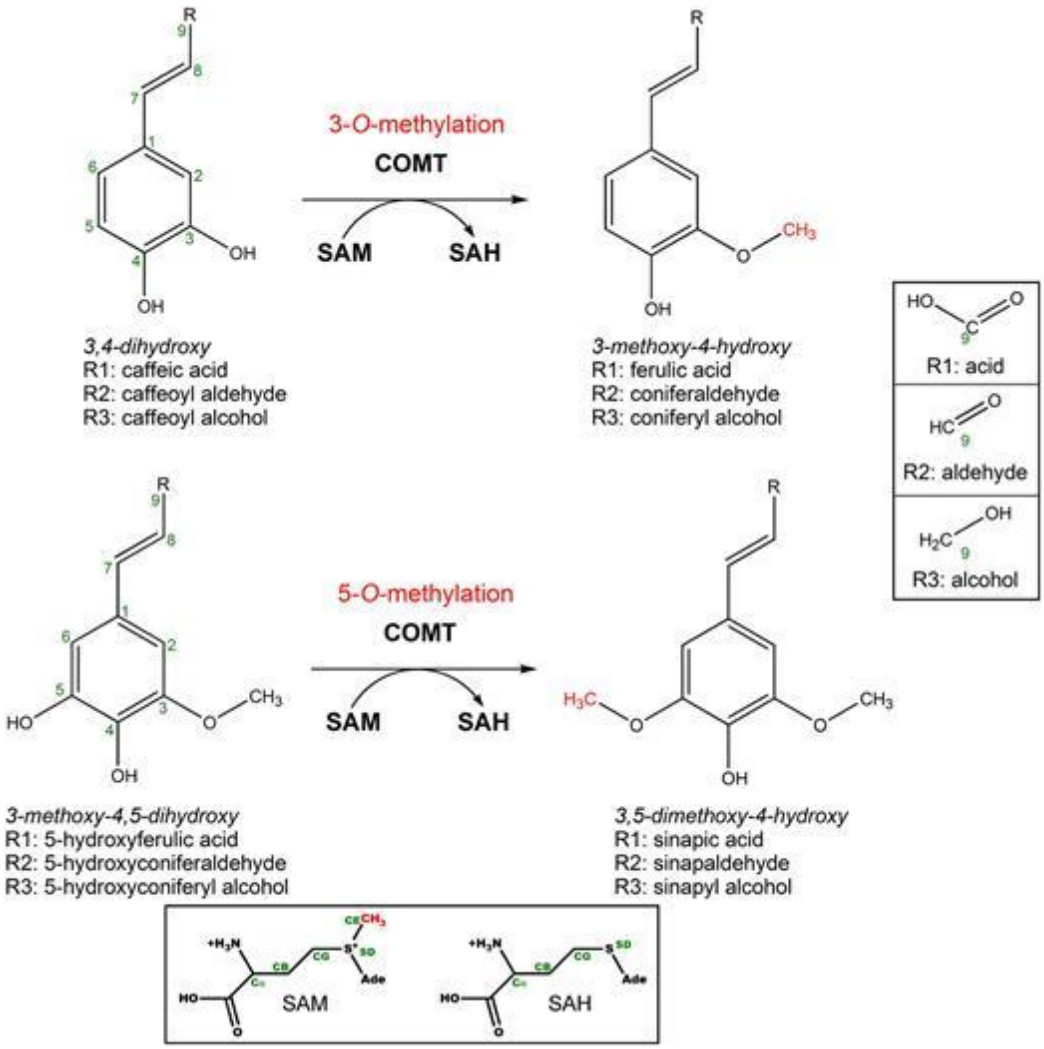


Figure 8: Potential Monolignol-Precursor Substrates of COMT and the *O*-Methylation Reaction Catalysed by COMT Enzymes (Louie et al., 2010)

Structure and active site

Crystal structures have been described for Lp OMT1 representing the holoenzyme as bound to S-adenosyl-L-homocysteine [SAH], and ternary complexes formed by interacting with SAH and sinapaldehyde or coniferaldehyde products. A functional binding pocket for the phenolic substrates like sinapaldehyde or coniferaldehyde is achieved by the inward positioning of the SAM binding domains associated with a closed conformational state of OMT1. The catalytic residues (His-266, Asp-267, and Glu-326) at the binding site and the SAM comes in close proximity of the phenolic substrate and SAM molecule traps the substrate in the binding pocket sequestering the phenoxide intermediate. In case of OMTs, it is observed that the methyl donor SAM binds first before the phenolic substrate and this clearly indicates the role of SAM in establishing the binding pocket (Huang et al., 2004) (Figure 9).

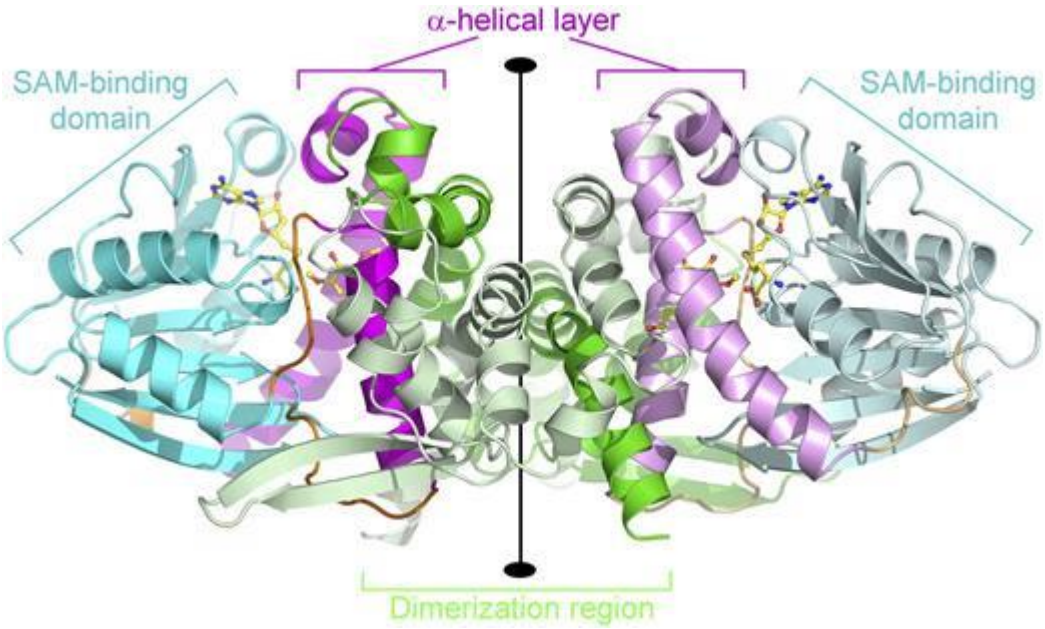


Figure 9: Structure of a Homodimer of the Lp OMT1/SAH/Sinapaldehyde Complex (Louie et al., 2010).

Caffeoyl CoA O-methyltransferase

Caffeoyl coenzyme A 3-O-methyltransferases (CCoAOMTs) are one important class of enzymes to carry out the transfer of the methyl group from S-adenosylmethionine to the hydroxyl group, and play important roles in phenylpropanoid biosynthesis (Ma and Luo, 2015).

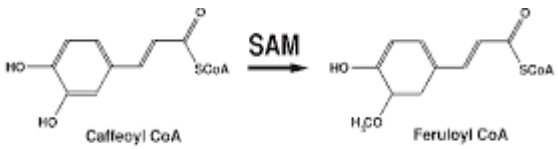


Figure 10: Caffeoyl CoA O-methyltransferase catalysed methylation reaction (Ferrer et al., 2005)

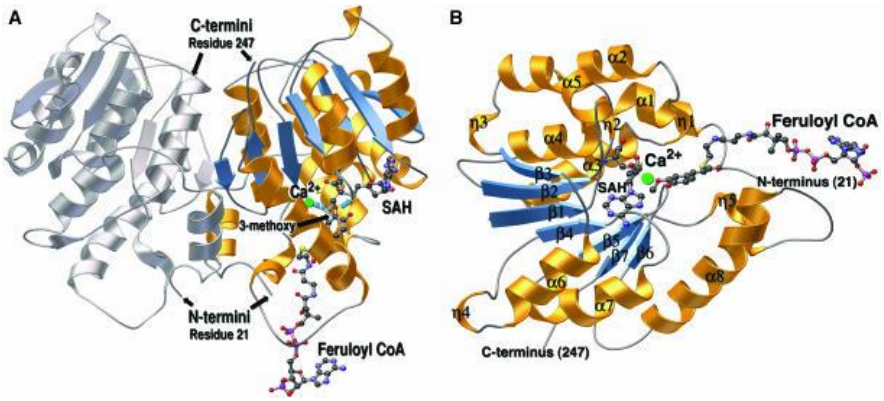


Figure 11: Ribbon diagrams of the CCoAOMT three-dimensional structure. **A**, Ribbon diagram of the alfalfa CCoAOMT homodimer, with one monomer colored (helices are gold and *β*-strands are blue) and the other monomer in grey. The SAH and feruloyl CoA molecules are depicted as carbon in black; nitrogen in blue; oxygen in red; and sulfur in yellow. The Ca^{2+} ion is in green. A blue arrow points from the same oxygen to the sulfur position on SAH has an attachment to the CH_3 group in SAM. The N-terminal and C-terminal amino acid residues in electron density map are numbered. **B**, Ribbon diagram of the CCoAOMT monomer. The feruloyl CoA and SAH molecules are depicted as sticks and ball and same colour codes as A are used.

Caffeoyl CoA 3-O-methyltransferase (CCoAOMT; EC 2.1.1.104) methylates coenzyme A esters and based on structure-function studies, a unique insertion in CCoAOMT among a variety of type 2 plant O-methyltransferases has been proposed as being responsible for CoA-thioester substrate specificity (Li et al., 1999). CCoAOMT are members of a multigene family (Lewis and Davin, 1999) and in parsley (Grimmig and Matern, 1997), grapevine (Busam et al., 1997) and alfalfa (Inoue et al., 1998) it has been reported to encode one to two members while in *Arabidopsis* (Raes et al., 2003) seven members are observed. CCoAOMT was first enzymes have been thought to use caffeoyl-CoA as a substrate and play important roles in the pathway of lignin biosynthesis. In many plants, caffeoyl-CoA O-methyltransferases (CCoAOMTs) were identified including *Medicago sativa* L., (Inoue et al., 1998), *Nicotiana tabacum* L (Martz et al., 1998), *Pinus taeda* (Li et al., 1999), *Populus* spp (Meyermans et al., 2000), *Mesembryanthemum crystallinum* L (Ibdah et al., 2003), *Lycopersicum esculentum* (Miao et al., 2008), *Acacia auriculiformis* x *Acacia mangium* (Pang et al., 2014), *Vitis vinifera* (Giordano et al., 2016), *Citrus reticulata* (Liu et al., 2020b) (Figure 11).

Structure and active site of caffeoyl CoA O-methyltransferase

Each monomer consists of a single catalytic domain. This domain exhibits a core α/β Rossmann fold that provides the binding site for SAM/SAH (Rossmann et al., 1974). All structurally characterized SAM-dependent OMTs share this architectural motif. The adenine portion of SAM/SAH forms hydrogen bonds with Asp-165 and the backbone nitrogen of Ala-140. Hydrogen bonds also form between the Rib hydroxyls and the carboxyl group of Asp-111. For SAH, the terminal-amino group hydrogen bonds with Gly-87 and Ser-93 and forms a salt bridge with Glu-85. The terminal carboxylate of SAH participates in a hydrogen-bonding network with Ser-93 and a well-ordered water molecule. Weaker electron density was observed in the active site for crystals soaked in the presence of caffeoyl CoA and 5-hydroxyferuloyl CoA. Active site is sequestered by a number of aromatic residues including Tyr-208, Tyr-212, and Trp-193, Met-61, Asp-163, and Asn-190 that aids in substrate binding (Ferrer et al., 2005).

There are several studies were conducted on O-methyltransferases involved in phenylpropanoid metabolism in anthocyanin biosynthesis, flavonoid biosynthesis, and lignin biosynthesis, alkaloid biosynthesis and stilbene biosynthesis. Both monomethylated and di-methylated anthocyanin were found in colored grape (*Vitis vinifera*) berries. A novel CCoAOMT was discovered in grapes that functioned as a bifunctional OMT both *in vitro* and *in vivo*, implying that this OMT alone may execute both methylation processes in grape berries (Hugueney et al., 2009). *Citrus reticulata* specific O-methyltransferases were identified which can methylate 3 ' -, 5 ' - and 7 ' - OH of flavonoids and further evaluated *in-vitro* cytotoxicity of methylated products showed that cytotoxicity of the flavonoids with the unsaturated C2-C3 bond was increased after being methylated at position 3 (Liu et al., 2020a). A novel multifunctional O-methyltransferase enzyme involved in lignin biosynthesis was identified from loblolly pine, which is a SAM-dependent enzyme that uses hydroxycinnamic acids hydroxycinnamoyl CoA Esters as substrate with both CAOMT and CCoAOMT activities and thus the potential to mediate a dual methylation pathway in lignin biosynthesis in loblolly pine xylem (Li et al., 1997). Several O-methyltransferase (OMT) enzymes, which participate in the terminal steps of monoterpene indole alkaloid biosynthesis, have been identified and cloned, including an OMT leading to the production of vindoline in *Catharanthus roseus* (Levac et al., 2008). An OMT producing ibogaine in *Tabernanthe iboga* (Farrow et al., 2018) and a

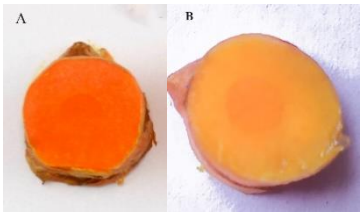
10-hydroxycamptothecin OMT from *Camptotheca acuminata* (Salim et al., 2018). A study involved the molecular characterization of O-methyltransferases involved in isoquinoline alkaloid biosynthesis in *Coptis japonica* and developed a chimeric OMT with novel substrate specificity from 6-OMT, 4-OMT, and SMT, and then expressed them in *E. coli* and characterized their reaction specificities (Morishige et al., 2010). Grapevine pterostilbene, a methyl ether of resveratrol, specific O-methyltransferase (ROMT) was identified through candidate gene approach and the activity of the corresponding protein was characterized after expression in *Escherichia coli*. The agroinfiltration approach was used to transiently coexpress ROMT and grapevine stilbene synthase in tobacco (*Nicotiana benthamiana*), resulting in the accumulation of stilbene in tobacco tissues (Schmidlin et al., 2008). For a deeper knowledge of the biosynthetic pathway at a molecular level, it is essential to use the high throughput sequencing technology. Comparative transcriptome analysis is the most effective strategy for identifying candidate genes involved in the biosynthetic pathway. The variations in curcumin content in turmeric accessions provide interesting material for identification of candidate genes (Deepa et al., 2016). A comparative transcriptome analysis of high papaverine mutant (pap1) and normal cultivar (BR086) of *P. somniferum*. Enhanced expression of (S)-norcoclaurine-6-O-methyltransferase (6OMT), (S)-3 ' -hydroxy-N-methylcoclaurine 4 ' - O-methyltransferase (4 ' OMT), norreticuline 7 ' -O-methyltransferase (N7OMT) and down-regulation of reticuline 7'-O-methyltransferase (7OMT) in pap1 in comparison to BR086 suggest (S)-coclaurine as the route for papaverine biosynthesis (Pathak et al., 2013). Comparative transcriptome analysis involving samples collected at various seasons in *Ampelopsis megalophylla* identified flavonoid biosynthesis genes. Accumulation of flavonoids was higher in May month than in the June and July months.

1.3. MATERIAL AND METHODS

1.3.1. Plant material

Transcriptome Sequencing and qRT-PCR: Rhizomes of IISR Prathibha, a released variety of ICAR-Indian Institute of Spices Research, Kozhikode (ICAR-IISR) collected from the Germplasm Repository (Experimental Farm, Peruvannamuzhi, Kozhikode, India), grown under rainfed conditions following the POP (Srinivasan et al., 2016) (PC- curcumin: >5.5%) was compared with that grown under irrigated and

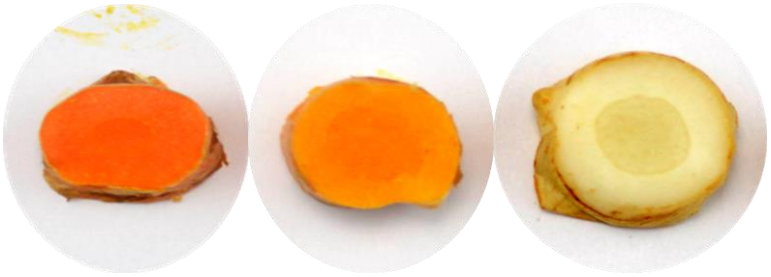
nutrient limiting conditions in a greenhouse (PS- curcumin- 1.5%) at ICAR-IISR Chelavoor farm, Kozhikode, India (11.298736408517987) (Figure 12).



PC PS

Figure 12: Cross section of the primary rhizomes used for transcriptome generation

Rhizomes of PC and PS were harvested at 120 DAP and were immediately transferred into RNA later solution (Sigma-Aldrich Chemicals Private Limited, Bangalore) and used for transcriptome generation. For qRT-PCR based expression of genes involved in curcumin biosynthesis IISR Prathibha, Acc.200 and *C. aromatica* were used (Figure 13).



(A) (B (C)

Figure 13: Cross section of (A) IISR Prathibha, (B) Acc.200 and (C) *C. aromatica*

Tissues such as rhizome, root, pseudostem, and leaves were collected from ICAR-IISR Prathibha, grown under field condition at ICAR-IISR, Farm, Chelavoor and was used for evaluating tissue specific expression (Figure 14). Rhizomes from four-five-month-old plants were used in all other experiments cited below. For evaluating impact of different geographical locations, IISR Prathibha, was planted in Erode, Tamil Nadu under irrigated condition compared with that grown under field condition and in green house at IISR Farm, Chelavoor (Figure 15).

Leaf Pseudo-stem Rhizome Root

Figure 14: Different tissues of *Curcuma longa* used for the study

Koz-field Koz-green house Erode-Field

Figure 15: *Curcuma longa* (turmeric) planted at three different geographical locations

For evaluating impact of different developmental stages, samples were collected at 120 DAP, 150 DAP and 180 DAP at IISR Farm, Chelavoor (Figure 16).

120 DAP 150 DAP 180 DAP

Figure: 16: Field view of different developmental stages of IISR Prathibha

To study the impact of different nutrient management systems on curcumin biosynthesis, ICAR-IISR Prathibha, grown under 100% Organic, Integrated and 100% Chemical management systems at IISR Experimental Farm, Peruvannamuzhi were used (Figure 17).

Figure 17: Study on impact of different management practices on curcumin biosynthesis. IISR Prathibha grown under 100% organic, integrated and 100% chemical management systems

To study the impact of different light regime, ICAR-IISR Prathibha, grown under red, green, white and black shade nets at IISR Experimental Farm, Peruvannamuzhi was used (Figure 18).

Figure 18: Study on impact of different light regimes on curcumin biosynthesis. IISR Prathibha grown under open, red, green, white and black shade net condition

1.3.2. Estimation of curcumin

The rhizomes were dried before transferring to RNAlater and used for curcumin quantification. The rhizomes were dried and ground to fine powder and curcumin content was analyzed spectrophotometrically at 430 nm following American Spice Trade Association (ASTA) procedure, 1968. All the samples used for this study were stored in RNAlater immediately after harvesting.

1.3.3. Total RNA extraction, cDNA library construction, and transcriptome sequencing.

Total RNAs were isolated using an optimized protocol described by Deepa et al., 2014. A260/280 and A260/230, RNA integrity number (RIN), and 28S/18S ribosomal RNA ratio were assayed using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA), and Agilent 2100 Bioanalyzer (Agilent Technologies, PaloAlto, CA, USA). Samples with RIN >7 and OD260/280 and OD260/230 values ≥1.8 were used for transcriptome sequencing. The transcriptome library was constructed using Illumina's TrueSeq RNA sample preparation kit as per the manufacturers' instructions and paired-end sequencing was generated using Illumina HiSeq^{TM} 2500 sequencing platform (Agri genome Labs Private Limited, Cochin, Kerala, India).

1.3.4. Transcriptome data processing and *de novo* assembly.

Fast QC (Andrews, 2010) was used to check the quality of the reads, and using Cutadapt (Martin, 2011) (https://github.com/marcelm/cutadapt) version 1.6 (Dortmund, Germany) adaptor sequences were removed, and other contaminations were removed using bowtie2 (version 2-2.2.3) (http://bowtie-bio.sourceforge.net/index.shtml), in-house Perl scripts and picard tools -version 1.85). Resulting the trimmed reads were then *de novo* assembled using short-reads assembly software Trinity (Grabherr et al., 2011) version trinityrnaseq_r20140717 (http://trinityrnaseq.sourcenforge.net) (with default options, minimum of 100 amino acid sequence) to obtain unigenes.

1.3.5. Functional annotation and classification of unigenes

Functional annotation of the assembled unigenes was performed by sequence comparison *via* BLASTX at default parameters against different public databases namely NCBI protein database (NR), NCBI nucleotide database (NT) UniProt (Bateman et al., 2021), (Bateman et al., 2015), KEGG (Kanehisa et al., 2016), (Dassanayake et al., 2009), TAIR(Huala et al., 2001), PDB(Berman et al., 2000), PROSITE (Hulo et al.,

2006)' (Sigrist et al., 2002), Pfam (Bateman et al., 2002), and PANTHER (Mi et al., 2005) and other databases. GO- and InterPro annotations were achieved using Blast2Go (Conesa et al., 2005) and InterProScan5 (Jones et al., 2014).

1.3.6. Transcriptome based differential gene expression analysis

Trimmed reads were aligned to the assembled transcriptome (length>=200bp) using Bowtie (Langmead, 2010). The expression level of each assembled transcript was measured using the fragments per kilobase per million mapped reads (FPKM) values. DEGs were identified by comparison between PC and PS using edgeR program following the criteria of minimum fold change ≥ ±1.00 and adjusted ρ-value of ≤0.05, with expression levels indicated as FPKM values. DEGs in case of PC and PS are identified and picturised as heatmaps. "R" program was used to classify the functional categories of molecular function, cellular component and biological process. KEGG pathway enrichment analyses were identified considering significantly enriched DEGs with a corrected P-value ≤0.05. Hierarchical clustering analysis was performed based on the log-transformed FPKM values using TBtool (Chen et al., 2020). DEGs identified between PC and PS were related to annotated genes involved in the curcumin biosynthesis pathway in *C. longa*.

1.3.7. qRT-PCR based differential gene expression analysis

To verify expression of candidate genes in transcriptomes, qRT-PCR was performed on nine selected genes and four selected transcription factors of the curcumin biosynthesis pathway. Total RNA used for qRT-PCR assays were the same as those used for transcriptome sequencing. The primers for the pathway genes were designed based on the contigs which were up-regulated in PC compared to PS, using the Primer quest tool (Primer Quest - design qPCR assays | IDT (idtdna.com)) following the criteria of GC% of 45–55% and melting temperature of 55–65 °C and size ~100 bp (Table 1).

Table 1: List of primers used in the study

Quantitative real time PCR primers for candidate reference genes		
	Forward primer (5'-3')	**Reverse primer (5'-3')**
EF1	CGGCTGCTCTGAGAAACAAT	GGTGATGCTGGACAAACAGTAG
UBIQUITIN	AAACCCAGTGGAGCAACTT	TATCGCTTGCGAGGCATATC
GAPDH	CAAGAGGAGCAAGACAGTTAGT	CAGATGCTCCTATGTTTGTGATTG
TUBULIN	GCGGAAGCAAATGTCGTAAAG	GGTAGAGCCATACAATGCTACTC
Quantitative real time PCR primers for genes		
PAL	TTGGATTGTTGTTGGTGGTTTAG	GAGGAGTTCGAGAAGGTGTTC
C4H	ATGACCGTGCCCTTCTTTAC	GGCGTCCTTCTTCACATCTT
4CL	GCTTTCGTCGTTCGGTCTAAT	GAACACCTTGCCTATCCTCTTG
HCT	CAGCTCATCCCAAAGGTAGAC	GTTGCATACCAACACCCAAAG
OMT	CTGCTCACCAAGCTCTCTATTC	TAGGCCTCTTCTTCGTCGT
CCoAOMT	GCGTCTACACGGGCTATTC	CAGCCCAATCTCGTAGTTCTC
DCS	GCTCGGACTGGAGAAGGA	CATCTCGTCCAGGATGAACAG
CURS1	ACCTTCCCGGACTTCTACTT	GCCTCTTCACCATCGTCTTC
CURS2	AAGGACATCGCCGAGAACAA	CTGGAAGTCGCGCTCGT
CURS3	GACCTTCCACTTCTTCAACCA	GTCGTTCCACTCGCTTATCC
ClPKS11	TGTCGGAGATCACCCACTTG	CGGAGAAGGGAGACCGAGA
Quantitative real time PCR primers for transcription factors		
bHLH	AGCTATGCGGCTTTGTACTC	CCCTTCAGCTCTCGCAAATA
WRKY	CCTTCATCGATCTCAACCTCTC	CTTTAATCTGCACCCGTCTTTG
bZIP	CTTTCTCCTCCAAGGCTTCTAC	GGATCAGAAGAGGCGATAGATTC
WD40	GTGTTCAGTGGTCTCCTGATAG	CCAGCATGTTCTTTCTGCTTAC
NAC	AAGCTCCTACCCTACGCTTAT	GTCGAGCATGCCTCCATTT
ERF	GGGCCAGGAAGAAGAAGATTAG	AAGCCGAACTCCATCAACTC
MADS box	CATGGCAAACACGGGAAAG	CGACGAAGCATTAGTCACAGA
HTH	TACACCATGGCCAATCACTC	CATCGGAGGCATCGTTCTT
Quantitative real time PCR primers for isoforms of Phenylalanine Ammonia Lyase		
***c45063_g3_i2**	TTGGATTGTTGTTGGTGGTTTAG	GAGGAGTTCGAGAAGGTGTTC
c45063_g3_i1	AAGCTGCGGATGGTGCT	CTTCGAAAGCCGCGATCTTATG
c45063_g2_i1	TGAATGCTTGAAGGAGTGGAATG	GCTCCTGATTTCATTGAACACAATTAC
c45063_g2_i6	CTGATCATCGATCCTCTGCTTG	CATTGGACACAATTCCACACTTT
c45063_g2_i8	CACGGTTTCTTCGAGCTGC	CGAGCAAAGCGAGGATGTT
c45063_g2_i4	TACATTGCAGGCCTTCTCAC	GCGAGCCCTTCCTTTGG
c45063_g2_i5	GGCGATCGGAAAGCTAATGT	CCCTTGAATCCATAGTCCAAACT
****ClPKS11***	TGTCGGAGATCACCCACTTG	CGGAGAAGGAGAGACCGAGA
Quantitative real time PCR primers for isoforms of O-methyltransferase		
****c42730_g1_i2**	CTGCTCACCAAGCTCTCTATTC	TAGGCCTCTTCTTCGTCGT
c41020_g1_i3	GCCAAGAACACGATGGAGAT	ACAACCCGATCCCGAATTAAG
*****c41020_g1_i2**	GCGTCTACACGGGCTATTC	CAGCCCAATCTCGTAGTTCTC
c41020_g1_i1	GAAGATGTAAACTCGATTCGACTAC	AGGAGCCGTGATTCTTCG
c32328_g1_i1	GCTGTGGCACTTCCTGAT	GATGGTCACGTTCCTCGTAAA
Full-length PCR Primers		
c42730_g1_i2	TTCGAGGTGATCATGACGTG	TGAGAAACCAGCAGCAAAAA
c41020_g1_i2	CCGACTCGAAGAATGGAGAG	GCTCGAGCAAGAAAAATGTTG

cDNA synthesis was performed using 250 ng of total RNA, first-strand Superscript III reverse transcriptase (Invitrogen) and Oligo- (dT)18 primer in a total volume of 20µl according to the manufacturer's instructions. Sample cDNAs were diluted to a final concentration of 100 ng/µL. qRT-PCR was performed in Qiagen Real Time PCR Cap Strip (Qiagen) on Rotor-Gene Q (Qiagen) using QuantiFast SYBR Green PCR kit (Qiagen). The reaction mixture included 10 µl of 2X SYBR Green, 10 pmol each of gene-specific primers, and 0.4 µl of cDNA in a final volume of 20 µl. Cycling conditions involved an initial denaturation at 95°C/5 min, followed by 40 cycles at 95°C/10s, and 60°C/30s. primer-specific annealing temperature was kept at 60°C/10s. A melt curve program of 65–99°C was included to check the specificity of PCR products. Three biological replicates were pooled together and each reaction had three technical replicates for analysing mean Ct value. To confirm the absence of genomic DNA and to check non-specific amplification, a reverse transcription negative control (without reverse transcriptase) and a non-template negative control were also included. Ubiquitin and EF1α genes served as internal reference genes and were used to normalize the gene expression data. The relative gene expression was calculated using the e $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2001) and transformed to log2 scale. A linear regression model was used to correlate the log-transformed relative quantification value of the genes from qRT-PCR results with the respective log-transformed relative gene expression values in the transcriptome data.

1.3.8. Cloning of full-length cDNA and sequence analysis of PAL and OMTs

Homologous transcripts to phenylalanine ammonia-lyase and O-methyltransferase were searched for within the text of annotated genes in transcriptome data of PC and PS. Default parameters of the Blast2GO software were used to retrieve gene ontology keywords for each unigene (Conesa and Gotz, 2008). RNA isolation, cDNA library preparation and qRT-PCR analysis were done as mentioned above. From the initial partial cDNA fragments obtained from PC, a full-length ClPAL2 cDNA clone was produced by 5′-3′ rapid amplification of cDNA ends (RACE). The RACE reactions were performed using the SMARTer™ RACE cDNA Amplification method by Clontech, USA. The primer sets are listed in Table 1. Genomic DNA was extracted from turmeric leaf using GenElute™ Plant Genomic DNA Miniprep Kit (Sigma) according to the manufacturer's instructions. After obtaining the full-length cDNA of ClPAL2 by RACE, primers were designed based on the sequences of the 5′-UTR and 3′-UTR, and

cDNA and genomic DNA were used as templates to clone the full-length ClPAL2 gene. The PCR conditions were as described by Li et al. (2012). The PCR products were separated by electrophoresis, and then the target gene fragments were purified using the QIAquick Gel Extraction Kit (Qiagen). The purified products were cloned into the pUC19 Linearized Vector and sequenced by the Agri genome (Kerala, India).

Deduced amino acid sequences of ClPAL2 was obtained by translating the cDNA sequences using the EMBOSS program TRANSEQ (Rice et al., 2000). Comparative sequence analysis of ClPAL2 was performed online using blastp (Altschul et al., 1990). The open reading frame (ORF) was predicted by ORF Finder (Rombel et al., 2002). Multiple sequence alignment of protein sequences with other PALs was conducted with ClustalW using default parameters. The phylogenetic analysis and construction of a neighbour-joining tree were performed with the MEGA 7.0 software using 1000 bootstrap replications.

1.3.9. Molecular modelling and docking

Structure homology was computed by the SWISS-MODEL (Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Switzerland) homology server (Waterhouse et al., 2018 and Bienert et al., 2017). The GMQE (Global Model Quality Estimation) and QMEAN parameter (Benkert et al., 2008) were determined to ascertain enough model quality. The model was validated using Ramachandran plot. Ramachandran plot was identified by Procheck program of Structural Analysis and Verification Server (Laskowski et al., 1993). The active site was predicted using COACH-D, an online accessible server (Wu et al., 2018). The canonical smileys of ligand molecule Phenylalanine [CC(C(=O)OC1=CC=CC=C1)N] are retrieved from the PubChem database (Kim, 2016; Kim et al., 2016a; Wang et al., 2017) and the 3D structure was created using CORINA (Sadowski et al., 1994; Sadowski and Gasteiger, 1993). Protein-ligand molecule interaction was studied using AutoDock 4.2 (Rizvi et al., 2013).

1.4. RESULTS AND DISCUSSION

1.4.1. Comparative transcriptome analysis and functional annotation

The Illumina HiSeqTM 2500 generated rhizome-specific transcriptomes of *Curcuma longa* under contrasting curcumin were employed for understanding the molecular process related to regulation of curcumin biosynthesis. Highest amounts of

curcumin were observed in rhizomes among all the different tissues. So, we have chosen rhizomes of a high curcumin accession, IISR Prathibha (Curcumin content >5.5%; PC) and same accession subjected to nutrient stress that showed a low curcumin content of 1.5%; PS. According to Huang et al., 2012a paired-end sequencing on the Illumina platform has been shown to increase the efficiency of *de novo* assembly and depth of sequencing. In non-model plants tissue specific transcriptome served as valuable resources for gene expression analysis (Zhou et al., 2012). Paired-end transcriptome sequencing generated 18.84 Gb of total raw reads for PC and PS. Similarly, 9.4 Gb of total raw reads were obtained in transcriptomes of *Gnetum Parvifolium* for mining candidate genes involved in flavonoids and stilbenoids (Deng et al., 2016). Average GC content of PC and PS was found to be 52.26 and 54.74 % respectively, indicating that the transcripts have almost equal AT and GC contents. This is comparable with the GC content of transcriptomes of *Lilium longiflorum* (Howlader et al., 2020). The raw reads generated from Illumina sequencing were deposited in National Centre for Biotechnology Information (NCBI), Sequence Read Archive database (SRA) under the Bio project ID PRJNA698442 (http://www.ncbi.nlm.nih.gov/sra). Summary of the sequencing outputs from PC and PS is shown in Table 2 and Table 3.

Table 2: Summary of the sequencing outputs of PC and PS transcriptomes

Sample Name	Total raw reads (Mb)	Number of paired-end reads	Total clean reads (Mb)	Total clean nucleotide (Gb)	GC (%)
PC	113.05	56.52	42.55	3.41	52.26
PS	75.30	37.65	51.73	4.13	54.74

Table 3: Summary of combined *de novo* transcriptome assembly of PC and PS

Description	Transcript length >=200bp
Number of assembled transcripts	138,015
Longest transcript length (bp)	37.043
Mean GC % of the transcript	46.94%

Overall, 99,834 unigenes (100%) could be successfully annotated with at least one significant hit in the NCBI database. Previous studies in *Curcuma longa* reported that only 56% of transcripts could be annotated (Annadurai et al., 2013). BLASTX homology search of RNA-seq data revealed that this might be due to the lack of genome information, poor quality reads and sequencing artefacts which lead to half of the transcripts in *Curcuma* to remain unannotated (Sheeja et al., 2015). In other crops such as *Brassica rapa* efficiency of alignment ranged from 86.20 to 89.54% (Wei et al., 2021), while in *Plumbago zeylanica* it is 80.71% (Sundari et al., 2020). However, only 39.05% of unigenes were annotated in *Dactylicapnos scandens* (He et al., 2016). Moreover, 77% of the assembled transcripts identified using BLASTX showed a similarity of more than 60% at the protein level with the existing proteins in the NCBI database. BLASTX similarity of more than 50% was observed in *Helianthus tuberosus* L when total unigenes were annotated using all databases (Zhao et al., 2020).

The species distribution map based on NR annotation revealed high similarities with *Musa acuminata subsp. Malaccensis* (66.06%), *Elaeis guineensis* (4.89%), *Phoenix dactylifera* (4.21%), *Ricinus communis* (2.41%), *Vitis vinifera* (1.03%), *Beta vulgaris subsp. vulgaris* (1.00%), *Oryza sativa subsp. indica* (0.83%) (Figure 19). More than 66% of transcripts showed identity with *Musa acuminata subsp. Malaccensis*, a member of the family *Zingiberaceae* which is related to *Curcuma*. This was followed by *Vitis vinifera, Oryza sativa etc*. as reported earlier (Sheeja et al., 2015). Clean reads for each sample were mapped back to the unigenes based on the assembly results, and the FPKM value of each unigene was calculated and used to measure gene expression levels. A total of 103,476 unique transcripts with FPKM ≥1 was identified, with highest number of expressed unigenes in PC; (86,949 with 88.02 alignment percentage) compared to PS (16527 with 84.34% alignment percentage).

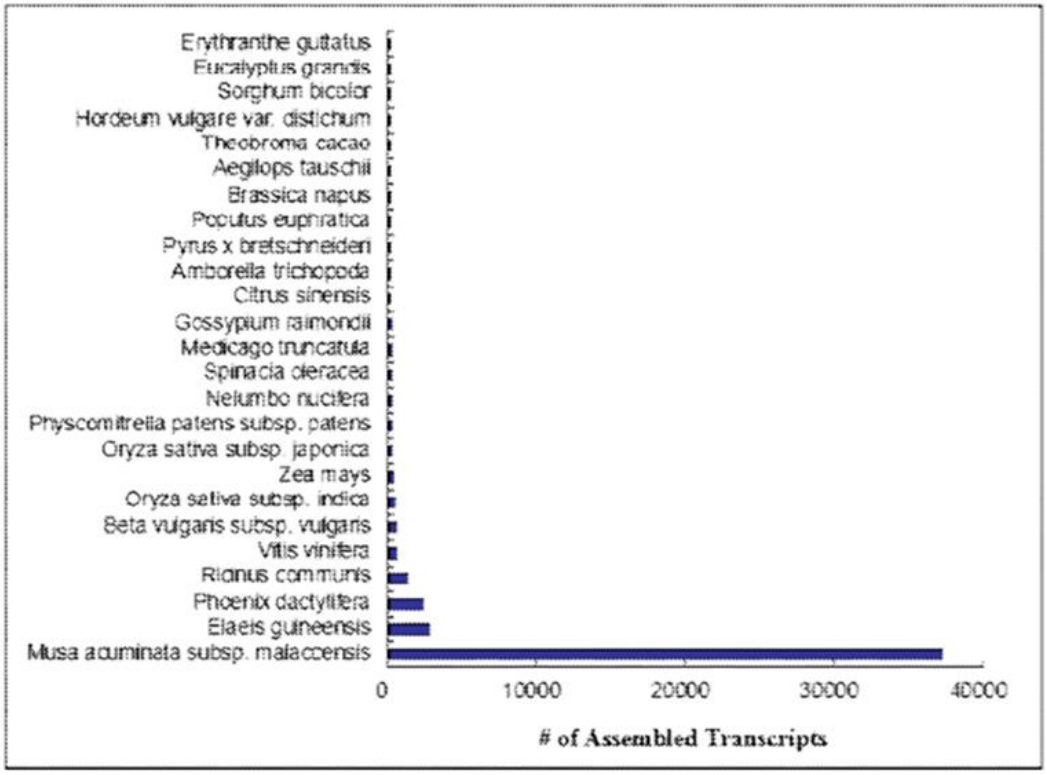
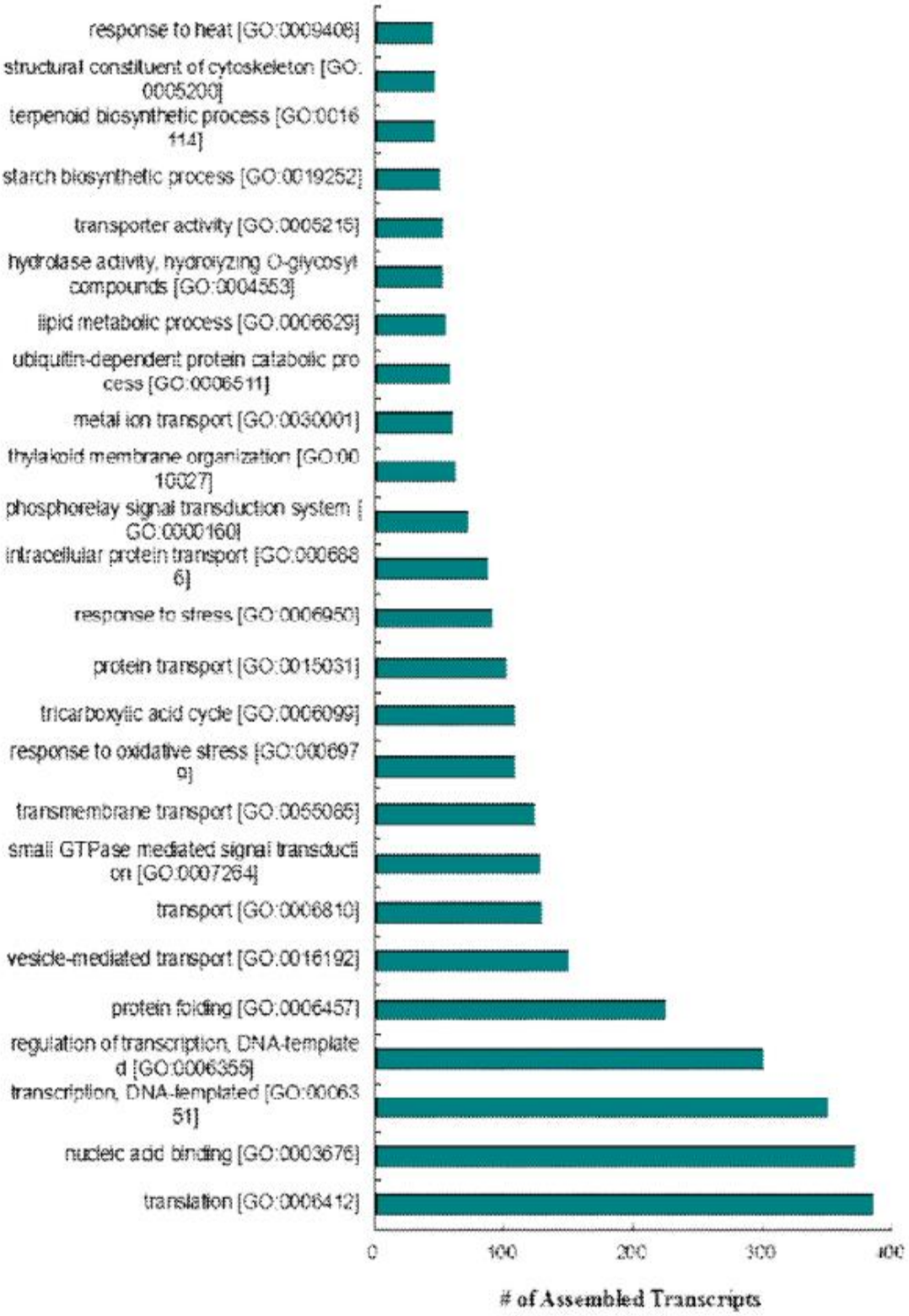


Figure 19: BLASTX top 25 hit organism distribution of transcriptome

Based on GO annotation, the unigenes were classifieds into categories of molecular function (3024) biological process (2402) and cellular component (1398). The terms ATP binding, translation and integral component of membrane represented the most represented among molecular function, biological process and cellular component categories, respectively. Top 25 terms of each category are represented in (Figure 20 (A) (B) (C)). The maximum transcripts fell into the category of translation, because translation plays a role in gene expression mechanism (Thomas et al., 2017) followed by nucleic acid binding being the ontology representation for transcription factors (Gaudet et al., 2021) . The subcategories *viz.,* 'secondary metabolic process' (GO:0019748), 'secondary metabolite biosynthetic process'(GO:0044550) and phenylpropanoid biosynthetic process (GO:0009699) contained 0.04%, 0.007% and 0.007% respectively. Similar GO term categories were reported in transcriptome study for mining candidate genes of phenylpropanoids in *Cryptomeria fortunei* Hooibrenk (Chinese cedar) (Yang et al., 2020). All the categories were downregulated in PS compared to PC.


calcium ion binding [GO:0005509]
integral component of membrane [GO:0016021]
magnesium ion binding [GO:0000287]
nucleotide binding [GO:0000166]
intracellular [GO:0005622]
DNA integration [GO:0015074]
sequence-specific DNA binding [GO:0043565]
heme binding [GO:0020037]
cytoplasm [GO:0005737]
hydrolase activity [GO:0016787]
protein serine/threonine kinase activity [GO:0004674]
ADP binding [GO:0043531]
catalytic activity [GO:0003824]
zinc ion binding [GO:0008270]
DNA integration [GO:0015074]
nucleus [GO:0005634]
transcription factor activity, sequence-specific DNA binding [GO:0003700]
protein kinase activity [GO:0004672]
structural constituent of ribosome [GO:0003735]
DNA binding [GO:0003677]
oxidoreductase activity [GO:0016491]
nucleic acid binding [GO:0003676]
metal ion binding [GO:0046872]
zinc ion binding [GO:0008270]
ATP binding [GO:0005524]
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
# of Assembled Transcripts


response to heat [GO:0009408]
structural constituent of cytoskeleton [GO:0005200]
terpenoid biosynthetic process [GO:0016114]
starch biosynthetic process [GO:0019252]
transporter activity [GO:0005215]
hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]
lipid metabolic process [GO:0006629]
ubiquitin-dependent protein catabolic process [GO:0006511]
metal ion transport [GO:0030001]
thylakoid membrane organization [GO:0010027]
phosphorelay signal transduction system [GO:0000160]
intracellular protein transport [GO:0006886]
response to stress [GO:0006950]
protein transport [GO:0015031]
tricarboxylic acid cycle [GO:0006099]
response to oxidative stress [GO:0006979]
transmembrane transport [GO:0055085]
small GTPase mediated signal transduction [GO:0007264]
transport [GO:0006810]
vesicle-mediated transport [GO:0016192]
protein folding [GO:0006457]
regulation of transcription, DNA-templated [GO:0006355]
transcription, DNA-templated [GO:0006351]
nucleic acid binding [GO:0003676]
translation [GO:0006412]
0
100
200
300
400
# of Assembled Transcripts

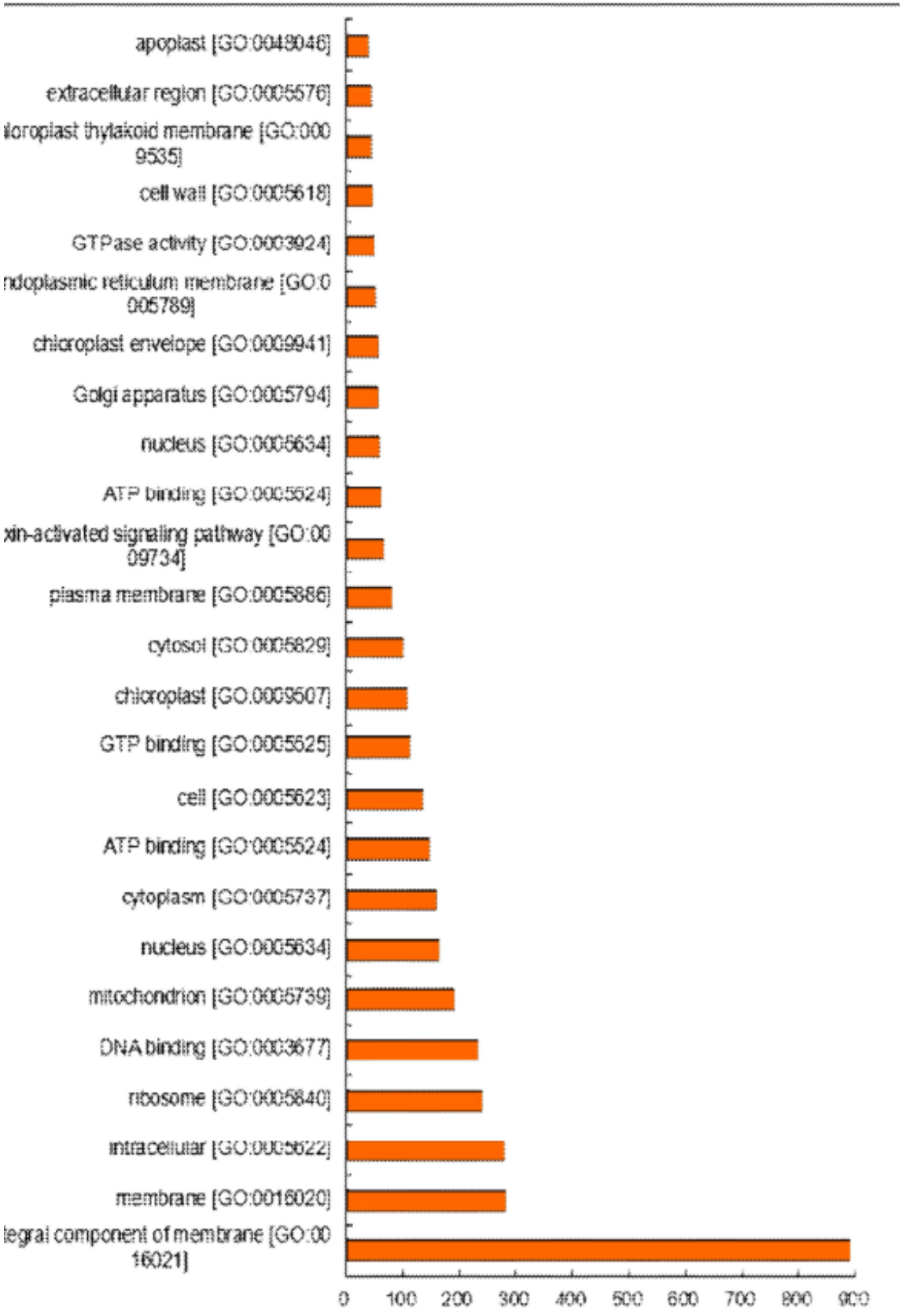


Figure 20: Top 25 terms in biological process, molecular function and cellular component from GO annotation

1.4.2. DEGs of PPP and validation of transcriptome data

Plants subjected to nutrient limiting conditions showed a lower curcumin content (1.5%), significantly different from the control (curcumin >5.5%). The nutrient stress lowered curcumin content as per our earlier observations. In related studies involving anthocyanin, salt stress induced accumulation of anthocyanin initially upto 2-4 days and a reduction was observed after 6-8 days (Chunthaburee et al., 2016).

Putative curcumin biosynthesis pathway was first proposed by Katsuyama (Katsuyama et al., 2009) in which phenylalanine is the first phenylpropanoid that is converted to p-coumaroyl CoA by, phenylalanine ammonia lyase (PAL), and later by cinnamate 4-hydroxylase (C4H) 3 and 4-coumarate CoA: ligase (4CL), the major upstream enzymes. Type III polyketide synthases that are downstream enzymes include diketide CoA synthase (DCS) and curcumin synthases (CURS 1-3) (Katsuyama et al. (2009 a, b). Another type III polyketide synthase named ClPKS10 putatively involved in curcumin biosynthesis was identified (Resmi and Soniya, 2012), however, not validated. Several pathway genes showing differential expression with respect to curcumin was identified through comparative transcriptome analysis of cultivated turmeric and a wild species *C. aromatica* (Sheeja et al., 2015) A novel type III polyketide, a downstream gene of the curcumin biosynthesis pathway was also identified (Deepa et al. 2017, Deepa, 2018).

In the present study a total of 103,476 DEGs could be detected using edgeR program (fold change $\geq \pm 1$ and adjusted p value ≤ 0.05) from individual pair of comparison between PC and PS. The highest number of DEGs observed in this case was 5,047 (5030 up and 17 down-regulated), and the lowest was between PS and PC (total 4913 with 4896 up and 17 down-regulated). The differential expression is depicted as fragment per kilobase of transcript per million mapped reads (FPKM) for each transcript/gene in each sample. The FPKM for each gene was calculated based on the length of the gene and read counts mapped to the gene (Fofana et al., 2020). Identification of differentially expressed genes is done mainly through fold change and FPKM values (Jiang et al., 2020). The transcript expression distribution is represented as FPKM (Figure 21).

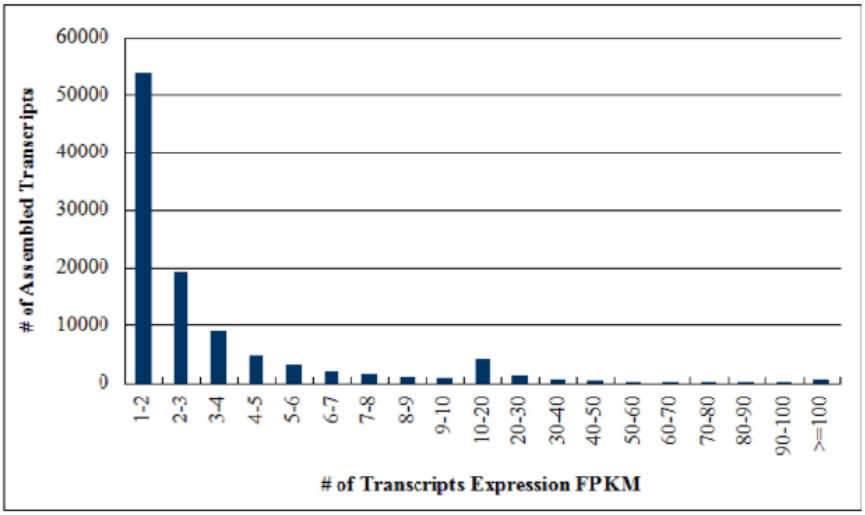


Figure 21: Distribution of transcript expression in PC and PS

More than 5000 transcripts showing FPKM between 1-2 could be observed in PC, while only 348 transcripts showing FPKM ≥ 100 was observed. Volcano plot represent the differentially expressed transcripts (Figure 22) in PC w.r.t PS. Each point represents transcripts and red-coloured points are upregulated genes and blue-coloured points represent downregulated transcripts, black points represent non-significant transcripts. Heat map based on fold change represent the complete set of upregulated and downregulated genes expressed in PC and PS (Figure 23 and 24).

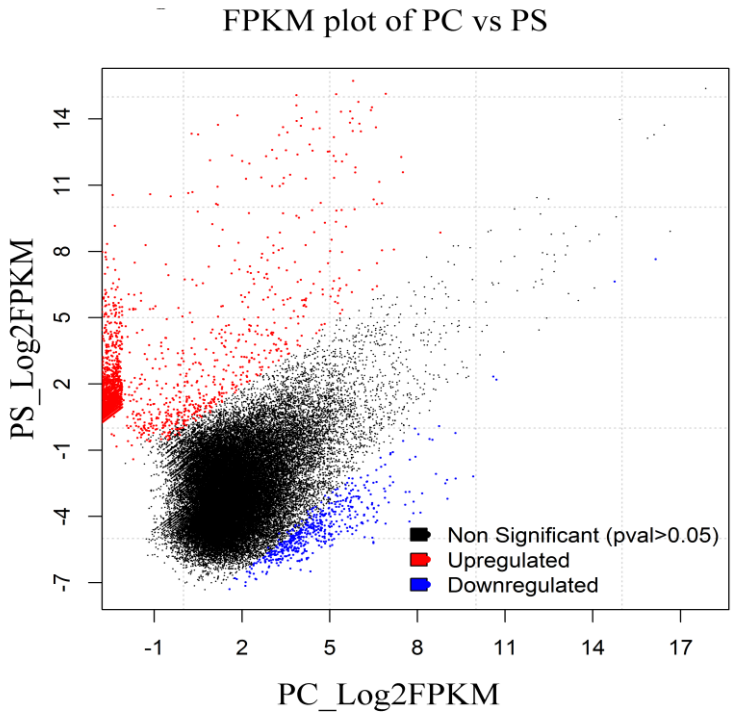


Figure 22: FPKM plot between PC and PS

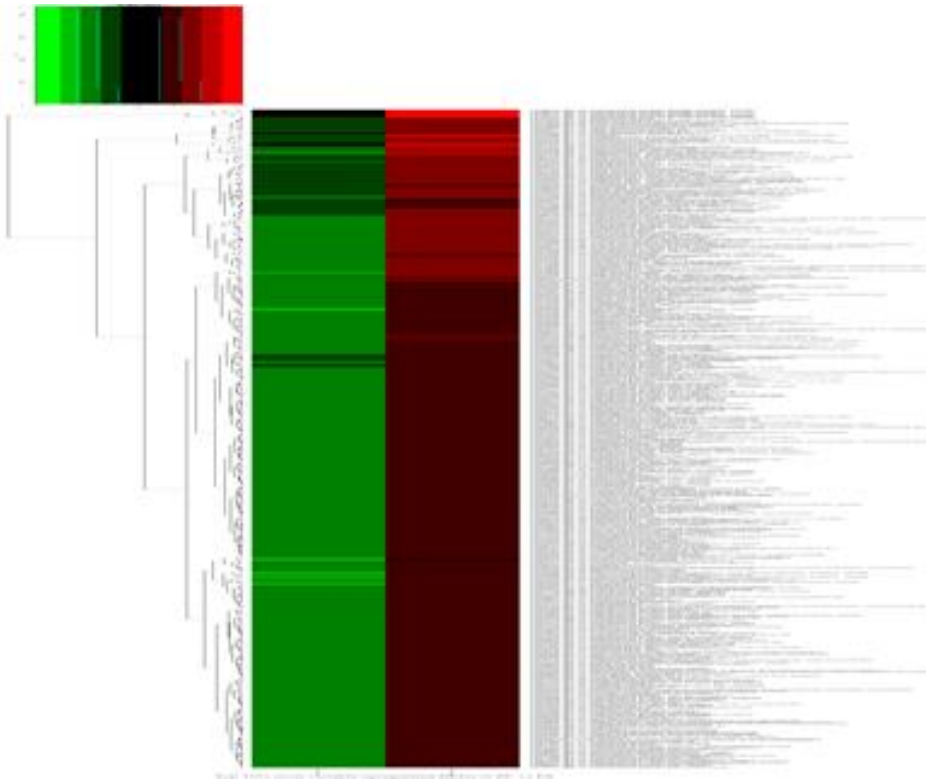


Figure 23: Top 10% most variable upregulated DEGs in PC vs PS

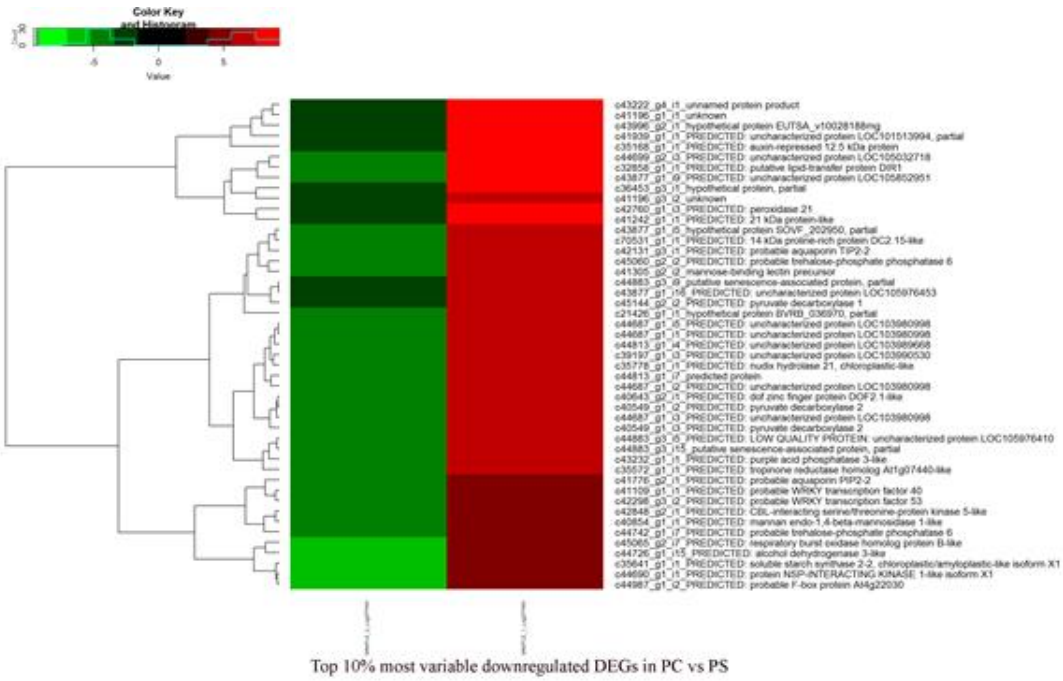


Figure 24: Top 10% most variable downregulated DEGs in PC vs PS

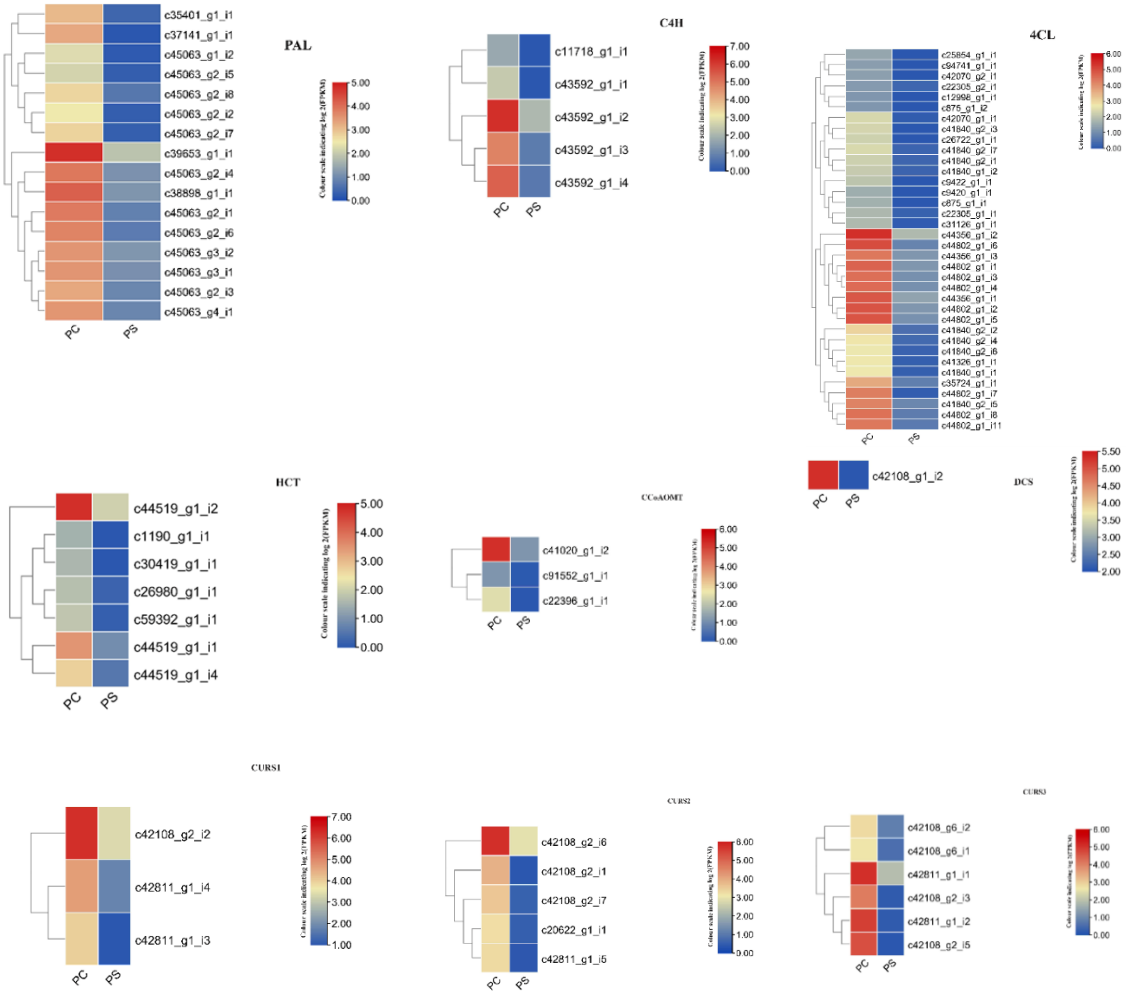


Figure 25: Heat map based on FPKM value of genes involved in curcumin biosynthesis obtained through comparative transcriptome analysis.

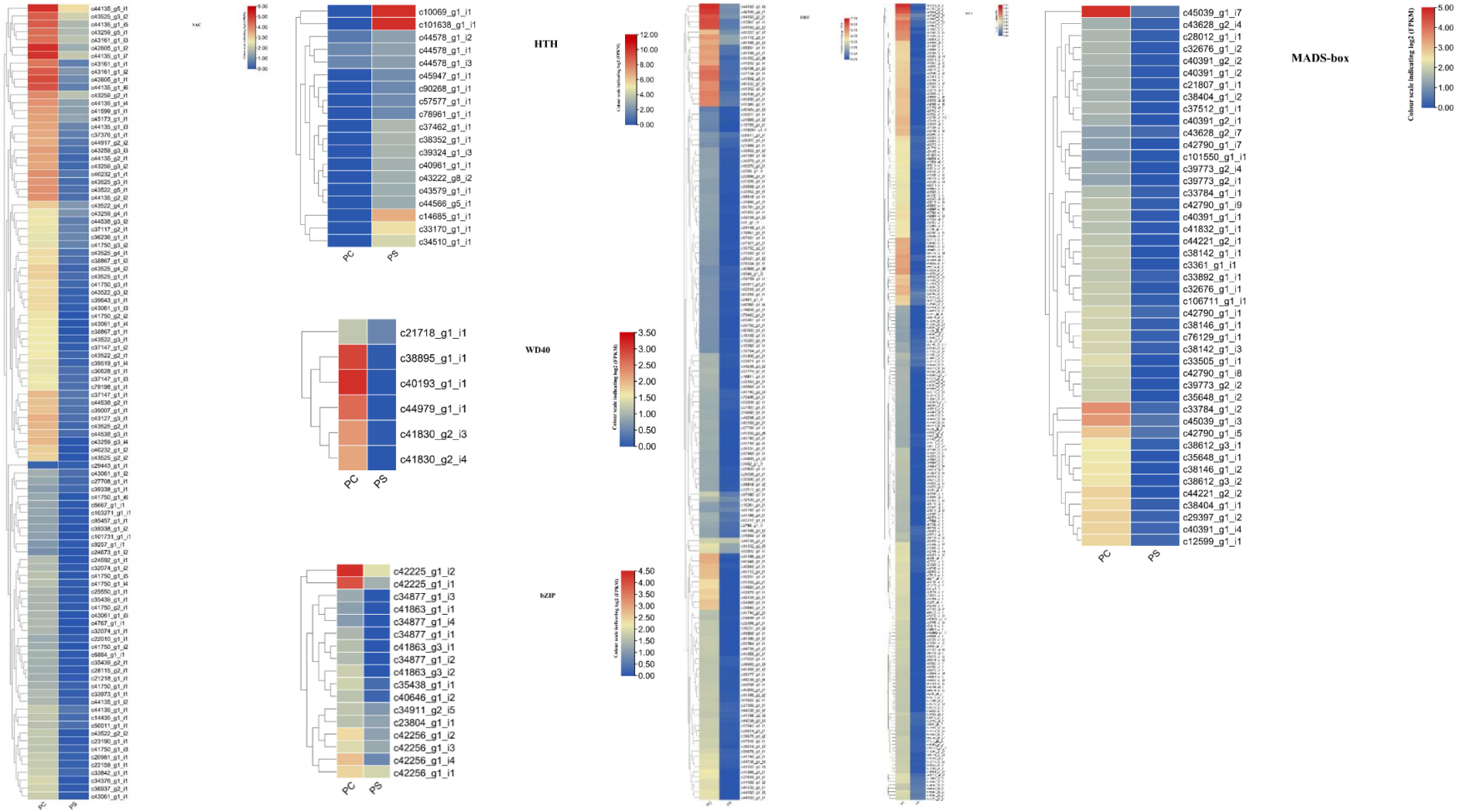


Figure 26: Heat map based on FPKM value of transcription factors involved in curcumin biosynthesis obtained through comparative transcriptome analysis.

In the present study, all the reported genes involved in the putative curcuminoid biosynthesis pathway could be detected from the rhizome-based transcriptome based on FPKM. PAL (16 unigenes), C4H (5 unigenes), 4CL (36 unigenes), HCT (7 unigenes), CCoAOMT (3 unigenes), DCS (1 unigene), CURS1 (3 unigenes), CURS2 (5 unigenes), CURS3 (6 unigenes), CLPKS11 (1 unigenes), OMT (27 unigenes) were present. And transcription factors such as bHLH (235 unigenes), WD40 (6 unigenes), bZIP (17 unigenes), WRKY (143 unigenes), NAC (101 unigenes) HTH (19 unigenes), MADS-box (45 unigenes), and ERF (155 unigenes) could be located in the RNA-seq data.

More than one unigene was annotated for the same enzyme, probably because they represent fragments of a single transcript, different members of a gene family or both (Hyun et al., 2012). Otherwise, most of the genes involved in PPP exist as multigene families and multiple isoforms (Bhuiyan et al., 2009; Lillo et al., 2008). Hierarchical clustering /heat map was done based on the FPKM values to select the candidates involved in curcumin biosynthesis (Figure 23). The results indicated that structural genes like *PAL, C4H, 4CL, HCT, CCoAOMT, DCS, CURS1, CURS2, CURS3, CLPKS11, OMT* and TFs like bHLH, WD40, WRKY, bZIP, HTH, MADS-box, ERF, and NAC showed differential expression with respect to curcumin in PS (curcumin- 1.52%) and PC (curcumin- >5.5%). (Figure 25 and Figure 26). This indicates that pathway genes exhibit differential expression *vis a vis* stress induced reduction in curcumin.

Comparative transcriptome is the best way to identify candidate genes involved in biosynthetic pathways. Several studies were done to mine the candidate genes of the pathway based on the differential expression of candidate genes *vis a vis* metabolite under conditions favouring differential accumulation of metabolites (Foong et al., 2020; He et al., 2018; Liu et al., 2015; Soltani Howyzeh et al., 2018). Validation of identified gene through qRT-PCR is mandatory. The present study involves dedicated pathway genes for curcumin biosynthesis and we have selected only important DEGs of the PPP leading to synthesis of curcumin. Eleven candidate genes and 8 TFs of PPP were selected for validation of transcriptome data by qRT-PCR. Among the four reference genes tubulin, GAPDH, elongation factor 1α (EF1α) were evaluated for transcript levels in PC and PS and to identify the most stable gene (Figure 27). Lowest Ct value was obtained for EF1α and showed less variability among samples. Analysis by different tools like Delta CT, Best Keeper, NormFinder and Genorm identified EF1α the most stable reference gene (Table 4).

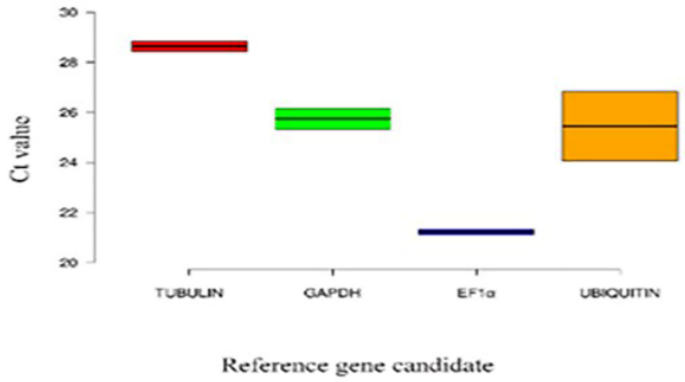


Figure 27: Expression of four candidate genes measured as Ct value by qRT-PCR and ranking order of stable reference gene predicted using RefFinder

Table 4. Ranking order of stable reference gene predicted using RefFinder

	Ranking order (Better-Good-Average)			
Method	1	2	3	4
Delta CT	EF1α	GAPDH	TUBULIN	UBIQUITIN
BestKeeper	EF1α	TUBULIN	GAPDH	UBIQUITIN
Normfinder	EF1α	GAPDH	TUBULIN	UBIQUITIN
Genorm	TUBULIN/EF1α		GAPDH	UBIQUITIN
Recommended comprehensive ranking	EF1α	TUBULIN	GAPDH	UBIQUITIN

Melt curve analysis of eleven structural genes and eight TFs by qRT-PCR after 40 cycles of amplification detected the presence of single peaks indicating that the only the expected amplicons were amplified in each case.

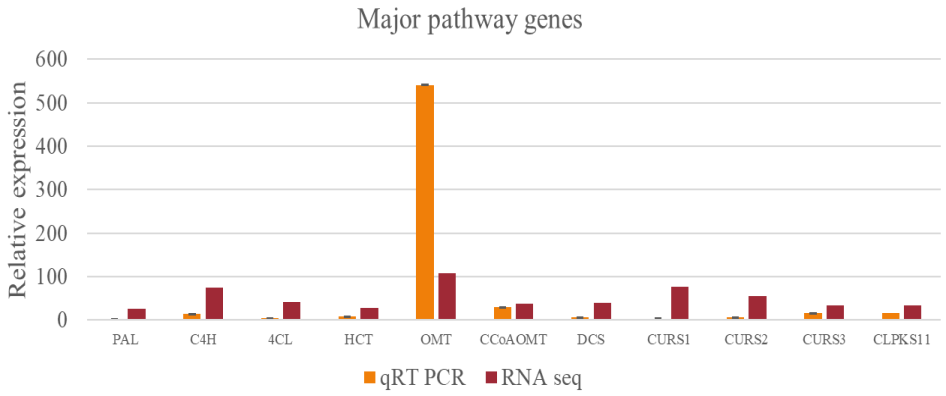


Figure 28: Differential expression of 11 selected pathway genes in PC (qRT-PCR and RNA-seq) with PS as control.

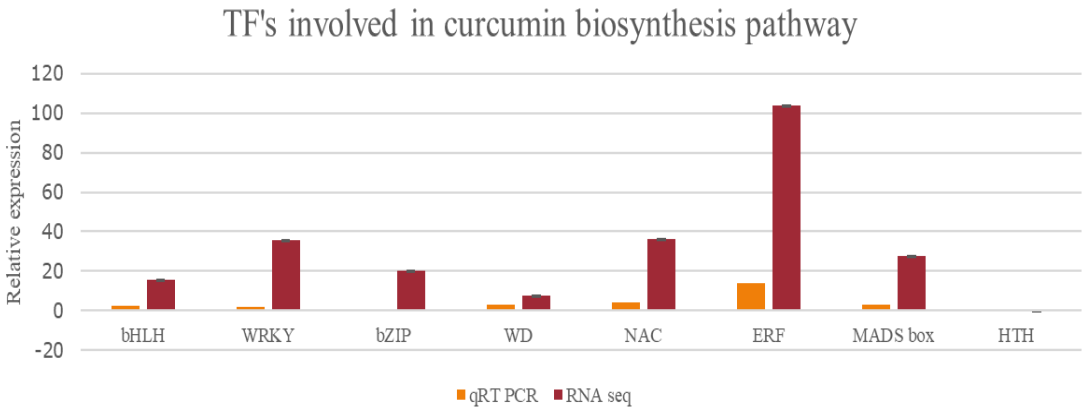


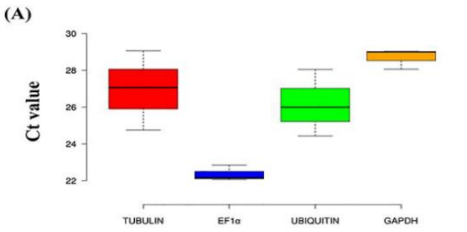
Figure 29: Differential expression of 8 selected curcumin biosynthesis pathway TFs in PC (qRT-PCR and RNA-seq) with PS as control.

Results of qRT-PCR data and transcriptome data correlated well suggesting that RNA-Seq data is reliable. Several studies have reported that RNA-seq data correlated well with qRT-PCR data (Xie et al., 2021: Tang et al., 2020). However, caffeoyl CoA O-methyltransferase gene showed much higher differential expression in qRT-PCR compared to transcriptome results, while rest of the genes and transcription factors (Figure 29) showed higher expression in the RNA-seq data. Several genes involved in MNQ biosynthesis in *Impatiens balsamina* L showed higher expression in qRT-PCR than RNA-seq (Foong et al., 2020)

1.4.3. Functional characterisation of key candidate genes- *PAL* and *OMT*

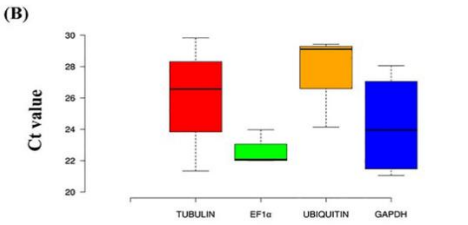
1.4.3.1. Identification of stable reference genes and expression analysis

Four reference genes viz., EF1 α, TUBULIN, GAPDH, ACTIN, were used for identification of stable gene for the experiments under spatio-temporal and tissue specific analysis along with different environmental and management practice were done. In all the experiments EF1α showing least variation among the samples was selected for normalizing the expression of target gene (Figure 30).



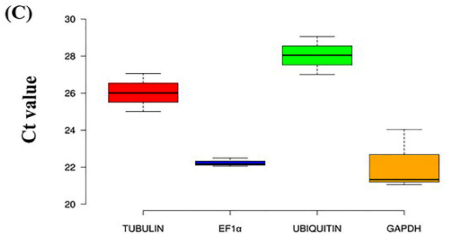
Ranking Order (Better--Good--Average)

Method	1	2	3	4
Delta CT	EF1α	GAPDH	UBIQUITIN	TUBULIN
BestKeeper	EF1α	GAPDH	UBIQUITIN	TUBULIN
Normfinder	EF1α	GAPDH	UBIQUITIN	TUBULIN
Genorm	EF1α | GAPDH		UBIQUITIN	TUBULIN
Recommended comprehensive ranking	**EF1α**	**GAPDH**	**UBIQUITIN**	**TUBULIN**



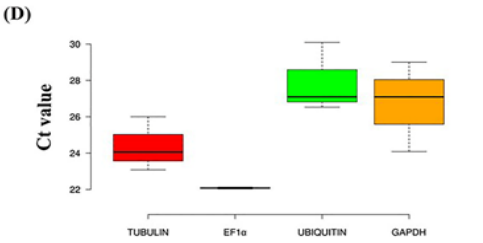
Ranking Order (Better--Good--Average)

Method	1	2	3	4
Delta CT	EF1α	UBIQUITIN	GAPDH	TUBULIN
BestKeeper	EF1α	UBIQUITIN	TUBULIN	GAPDH
Normfinder	EF1α	GAPDH	UBIQUITIN	TUBULIN
Genorm	EF1α | UBIQUITIN		GAPDH	TUBULIN
Recommended comprehensive ranking	**EF1α**	**UBIQUITIN**	**GAPDH**	**TUBULIN**



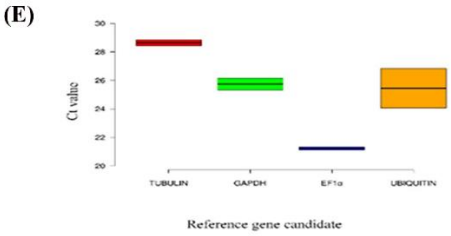
Ranking Order (Better--Good--Average)

Method	1	2	3	4
Delta CT	EF1α	TUBULIN	GAPDH	UBIQUITIN
BestKeeper	EF1α	UBIQUITIN	TUBULIN	GAPDH
Normfinder	EF1α	TUBULIN	GAPDH	UBIQUITIN
Genorm	TUBULIN | GAPDH		EF1α	UBIQUITIN
Recommended comprehensive ranking	**EF1α**	**TUBULIN**	**GAPDH**	**UBIQUITIN**



Ranking Order (Better--Good--Average)

Method	1	2	3	4
Delta CT	EF1α	TUBULIN	UBIQUITIN	GAPDH
BestKeeper	EF1α	TUBULIN	UBIQUITIN	GAPDH
Normfinder	EF1α	TUBULIN	UBIQUITIN	GAPDH
Genorm	TUBULIN | UBIQUITIN		EF1α	GAPDH
Recommended comprehensive ranking	**EF1α**	**TUBULIN**	**UBIQUITIN**	**GAPDH**



Ranking Order (Better--Good--Average)

Method	1	2	3	4
Delta CT	EF1α	GAPDH	TUBULIN	UBIQUITIN
BestKeeper	EF1α	TUBULIN	GAPDH	UBIQUITIN
Normfinder	EF1α	GAPDH	TUBULIN	UBIQUITIN
Genorm	TUBULIN | EF1α		GAPDH	UBIQUITIN
Recommended comprehensive ranking	**EF1α**	**TUBULIN**	**GAPDH**	**UBIQUITIN**

Figure 30: Expression of four candidate genes by qRT-PCR measured as Ct values and ranking order of stable reference gene predicted using RefFinder.(A,B,C,D, and E are spatio-temporal, tissue specifi, different management practices and different light regime respectively)

1.4.3.2. Shortlisting of candidate genes and isoforms of *PAL* from transcriptome

Among the genes and transcription factors analysed in the previous section, we have chosen only important candidate genes for further elaborate studies due to limitations in resource and time. Phenylalanine ammonia lyase and O-methyl transferase are two important genes of the curcumin pathway and *PAL* is being considered as the rate limiting gene. Through differential gene expression analysis of transcriptomes of PC and PS *vis a vis* curcumin, seven *PAL* gene transcripts involved in curcuminoid biosynthesis could be identified in our study with minor differences in their sequences (Table 5).

Table 5: Gene ontology analysis and blast description of phenylalanine ammonia lyase

Contig ID	FPKM value in PC	FPKM value in PS	Sequence description	Blast top hit description
c45063_g3_i2	10.61	1.24	*Curcuma longa* phenylalanine ammonia lyase (PAL) mRNA, partial cds	KJ780359.1
c45063_g3_i1	10.60	1.12	*Curcuma longa* phenylalanine ammonia lyase (PAL) mRNA, partial cds	KJ780359.1
c45063_g2_i8	5.73	0.59	*Curcuma alismatifolia* phenylalanine ammonia lyase 1 mRNA, complete cds.	MT811933.1
c45063_g2_i4	13.95	1.16	*Hedychium coronarium* phenylalanine ammonia lyase (PAL) mRNA, complete cds	MW415433.1
c45063_g2_i1	13.58	0.79	*Curcuma longa* phenylalanine ammonia lyase (PAL) mRNA, partial cds	KJ780359.1
c45063_g2_i6	12.35	0.68	*Curcuma longa* phenylalanine ammonia lyase (PAL) mRNA, partial cds	KJ780359.1
c45063_g2_i5	3.46	0.22	*Curcuma longa* phenylalanine ammonia lyase (PAL) mRNA, partial cds	KJ780359.1

Many studies have reported identification of candidate *PAL* genes associated with phenylpropanoid biosynthesis in crops like *L. chinense* (Zhao et al., 2013), *O. saundersiae* (Wang et al., 2014), following a transcriptomic approach. *PAL* isoforms with minor difference in 3' UTR was reported in *Phaseolus vulgaris* (Cramer et al., 1989). Several mechanisms like alternative splicing, intron retention, and alternative transcription start/stop sites help to diversify mRNA sequences, yielding isoforms that often differ in their protein-coding capacity and function (Ray et al., 2020). Seven isoforms of PAL such as c45063_g3_i2, c45063_g3_i1, c45063_g2_i8, c45063_g2_i4, c45063_g2_i1, c45063_g2_i6, and c45063_g2_i5 had sequence lengths of 961 bp, 931 bp, 778 bp, 697 bp, 1124 bp, 1134bp, and 334 bp respectively. Previous studies showed that in most of the cases, PAL full length sequences are larger in size above 2000 bp (Jin et al., 2013; Ma et al., 2013; Zhu et al., 2015). Short read lengths are a critical barrier to accurate assembly of highly similar gene isoforms in Next-generation sequencing technologies like Illumina sequencing platform (Liang et al., 2016). Among those seven isoforms maximum FPKM was observed in c45063_g2¬i4 (13.95 in PC and 1.16 in PS) and minimum in c45063_g2_i5 (3.46 in PC and 0.22 in PS). All those seven isoforms were up regulated in IISR Prathibha (>5.5% curcumin), under field conditions and showed >85% similarity with the already reported plant PALs. Characterization of the full complement of isoforms within targeted genes or across an entire genome is technically challenging due to the high degree of sequence identity among these isoforms (Liang et al., 2016). The transcript c45063_g3_i2 showed 92.04% similarity with *Curcuma longa* phenylalanine ammonia-lyase (PAL) mRNA, partial cds (KJ780359.1). Similarly, four isoforms namely c45063_g3_i1, c45063_g2_i1, c45063_g2_i6, c45063_g2_i5 showed similarity with already reported *Curcuma longa* phenylalanine ammonia-lyase (PAL) mRNA, partial cds (KJ780359.1) with percentage similarities of 91.77%, 97.51%, 99.12% and 88.20% respectively. Other isoforms such as c45063_g2_i8 and c45063_g2_i4 64.62% showed similarity with *Curcuma alismatifolia* phenylalanine ammonia-lyase 1 mRNA, complete cds (MT811933.1) and *Hedychium coronarium* phenylalanine ammonia-lyase (PAL) mRNA, complete cds (MW415433.1) with 93.53% similarity. *Hedychium coronarium* also called white butterfly ginger lily belongs to the Zingiberacae family and four *PAL* isoforms *viz.*, c45063_g3_i2,

c45063_g3_i1, c45063_g2_i1 and, c45063_g2_i6 showed >90% identity with *Curcuma longa* phenylalanine ammonia-lyase (PAL) mRNA, partial cds (KJ780359.1). This suggests that these contigs might be members of multigene families or part of the same gene without overlapping sequences that form single large contigs or may be involved in alternate biosynthetic pathways other than curcumin (Upadhyay et al., 2014).

1.4.3.3. Expression analysis of selected isoforms of *PAL*

The expression patterns of seven *PAL* isoforms were studied in 4-month-old rhizomes of IISR Prathibha (>5.5% curcumin), Acc 200 (3.50% curcumin) and *C. aromatica* (0.12% curcumin). *ClPKS11* were used as bait gene (Deepa et al., 2017) and *C. aromatica* was kept as control. Among those transcripts c45063_g3_i2 showed enhanced expression (fold change of 126 and 5.2 in high and low curcumin accessions respectively) in high curcumin accession and *vice versa* similar to the expression of the bait gene, *ClPKS11* (fold change-20 in high curcumin accession; 0.7 in low curcumin accession). C45063_g3_i1 (fold change -2 in high curcumin accession; 1.0 in low curcumin accession) and c45063_g2_i1 (fold change-1.3 in high curcumin accession; 0.5 in low curcumin accession) also showed a similar pattern, though the expression was low. Expression profiles of c45063_g2_i6 (fold change-0.9 in high curcumin accession;0.4 in low curcumin accession) and c45063_g2_i5 (fold change-0.4 in high curcumin accession;0.17 in low curcumin accession) and c45063_g2_i4 (fold change-3.45 in high curcumin accession; 3.07 in low curcumin accession) did not show any correlation with curcumin, while c45063_g2_i8 (fold change-0.7 in high curcumin accession; 3.68 in low curcumin accession) showed a negative correlation with respect to curcumin (Figure 31 (A)). The expression results were analyzed using the comparative cycle threshold (Ct) method and quantified relative to the control ($2^{-\Delta\Delta Ct}$) (Schmittgen and Livak, 2001). qRT-PCR based identification and classification of isoforms is a reliable and robust method already reported in plants (Harvey et al., 2016; Vandenbroucke et al., 2001)

Influence of different tissues in determining curcumin status is still unknown (Ayer et al., 2020). Expression patterns of these isoforms in different tissues such as leaves, pseudostem, root, and rhizomes of IISR Prathibha planted in field at Kozhikode

120 DAP was evaluated using pseudo stem as control. C45063_g3_i2 was upregulated in the rhizome (fold change-21.70) and down regulated in both leaf (fold change-0.02) and root (fold change-0.05). Results indicated that expression of C45063_g3_i2 and curcumin is high in rhizome compared to leaf and root similar to reports by Behar et al., 2016. However, expression of c45063_g3_i1 (fold change-0.45) and c45063_g2_i8 (fold change-0.52) was high in leaf followed by rhizome (fold change-0.20 and 0.41 respectively) with least expression observed in the root (fold change-0.01 and 0.03 respectively). Expression of c45063_g2_i1(fold change-0.05) and c45063_g2_i4 (fold change-0.37) was high in rhizome followed by root (fold change-0.03 and 0.17 respectively) and leaf (fold change-0.01 and 0.04 respectively). It is presumed that these isoforms perform alternate functions in leaf, which is an interesting matter for detailed study. In case of c45063_g2_i5, the highest expression was observed in the root (fold change-0.26), followed by rhizome (fold change-0.168) and leaf (fold change-0.08) (Figure 31 (B)). Studies conducted in different crops on tissue specific expression of *PAL* revealed that *PAL* gene performs are involved in multiple physiological roles in different tissues (Hashemitabar et al., 2014; Jin et al., 2013; Xu et al., 2012). Specific metabolites often show a restricted presence in few organs, tissues, and cell types (Li et al., 2016), and can be extensively regulated by environmental factors (Li et al., 2015; Tohge et al., 2016).

The contig C45063_g3_i2, a novel transcript that correlated well with curcumin in different tissues and varieties was designated as *ClPAL2* and was chosen for further studies involving qRT-PCR. Evaluation at different geographical locations with IISR Prathibha as control indicated a positive correlation with curcumin when planted at three different environmental conditions. From the study it was observed that IISR Prathibha grown in Kozhikode field condition (Koz-field) showed highest ClPAL2 expression (fold change-2) compared to Koz-green house (fold change-0.24) and Erode-field (fold change-0.03) (Figure 31 (C)). The curcumin content of IISR Prathibha was higher under field condition at Kozhikode (≥5.5%) than Erode-field (3.68%) than Koz-green house (3.01%). It is evident that curcumin is subjected to environmental variations and conditions of growth. It has been reported that curcumin content varies with the different accessions (Sasikumar, 2005b, Sajitha et al., 2014) and subjected to agroclimatic

variations (Anandaraj et al., 2014; Singh et al., 2013). Our results suggested that cultivation under rainfed condition at Kozhikode is the most ideal condition for growth that favoured the production of curcumin in IISR Prathibha. Similarly, in many plants many workers have reported that environmental factors had an effect on anthocyanin biosynthesis (Annis et al., 2019; Chalker-Scott, 1999; Piccaglia et al., 2002). The accumulation of anthocyanins depends on both low temperature and light, and these factors have a synergistic effect on the expression of genes within the anthocyanin biosynthesis pathway (Azuma, 2018).

Expression profiling was done under different developmental stages (120 DAP, 150 DAP, and 180 DAP) and in different varieties (IISR Prathibha, Alleppey supreme, Acc. 200, *C. aromatica* and *C. ceasia*). The results indicated highest expression (fold change-99.73) in IISR Prathibha at 120 DAP. Expression drastically reduced (fold change-4.78) at 150 DAP with further reduction at 180 DAP (fold change- 3.03). Expression of *ClPAL2* was down-regulated in both *C. aromatica* (0.12% curcumin) (fold change-0.08, 0.22, and 0.03 respectively for 120 DAP, 150 DAP and 180 DAP) and *C. ceasia* (0.16% curcumin) ((fold change-0.14, 0.40, and 0.22 respectively for 120 DAP, 150 DAP and 180 DAP)) at all developmental stages (Fig 2c), which correlates with the low curcumin in these related species. Results showed that expression of ClPAL2 was maximum at 120 DAP *vis-à-vis* curcumin content and reduced at 180 DAP. The result is consistent with the biochemical study and this might be attributed to the reason that curcumin levels are more or less stable after 180 DAP (Neema, 2005). Similarly temporal expression patterns of phenylalanine ammonia lyase were studied in *Rubus idaeus* which correlated well with the amount of metabolite under different developmental stages (Kumar and Ellis, 2001). Evaluation in different accessions indicated a positive correlation with curcumin content (Figure 31 (D)), with maximum in IISR Prathibha (>5.5% curcumin; fold change-133.43) followed by Alleppey supreme (5.21% curcumin; fold change-6.1) with lowest expression in *C. aromatica* (0.12% curcumin; fold change-0.01).

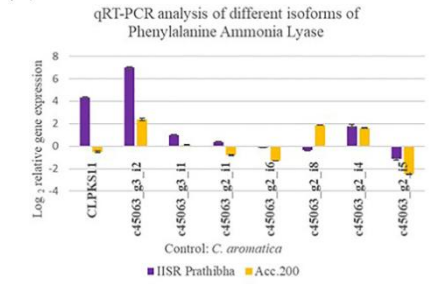
Evaluation of gene expression under different management practices indicated that integrated management showed highest expression of *ClPAL2 (*fold change-7.67*)*

while 100% chemical system showed lowest expression (fold change-0.25) (Figure 31 (E)). Results indicated the importance of integrated/organic management systems over chemical management for maximizing gene expression and curcumin content. Similar results were obtained for ClPKS11 where the expression and curcumin content were high under integrated management systems than 100% organic and 100% chemical systems.

The influence of different shade nets such as red, white, green, and black on gene expression and curcumin accumulation in IISR Prathibha was evaluated with open condition as control. The highest expression was observed in plants grown under green shade net (fold change-2.49: 4.83% curcumin) followed by black (fold change-0.51; 4.59% curcumin), white (fold change-0.27; 4.30% curcumin) and red (fold change-0.125:4.02% curcumin) (Figure 31 (F)). Gene expression studies of ClPAL2 indicated that expression was reduced in red, white and black shade net conditions while green shade net gave maximum expression. This indicates the influence on light regimes on the expression of ClPAL2 and thereby curcumin levels. Similarly in leaves and calli of crab apple (*Malus* spp) anthocyanin and anthocyanin biosynthesis genes were decreased leading to a reduction in colour of leaf and calli (Lu et al., 2015). Foliar anthocyanins and their relative abundance was investigated in leaves of sweet potato under shading and a reduction in expression of anthocyanin biosynthesis genes and anthocyanin content was observed (Islam et al., 2005).

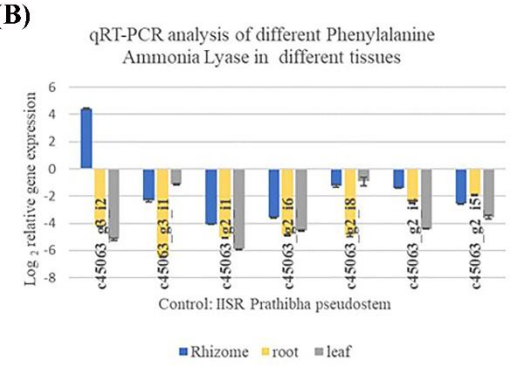
To sum up, among the seven *PAL* isoforms analyzed, expression of *ClPAL2* showed positive correlation with curcumin content under all the different experiments involving developmental stages, tissues, light regimes, geographical location and management practices. The findings thus indicate that *ClPAL2* could be a novel gene involved in regulation of curcumin biosynthesis in turmeric and it is prospective to utilise intrinsic levels of *ClPAL2* as a marker for screening for curcumin.

(A)



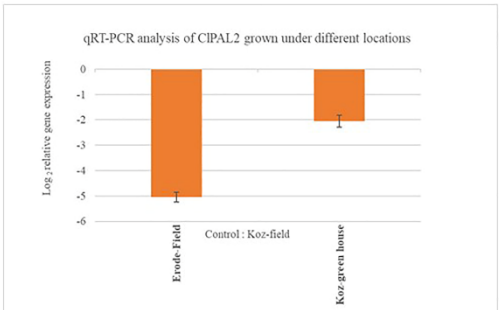
Different turmeric accessions	Curcumin content(%)
IISR Prathibha	5.37
Acc. 200.	3.50
C. aromatica	0.12

(B)



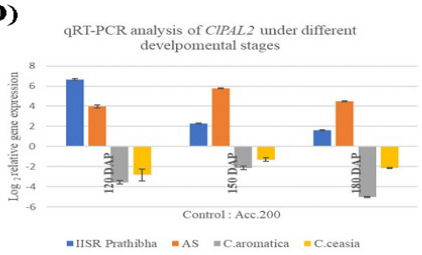
Different locations	Curcumin content(%)
IISR Prathibha rhizome	5.37

(C)



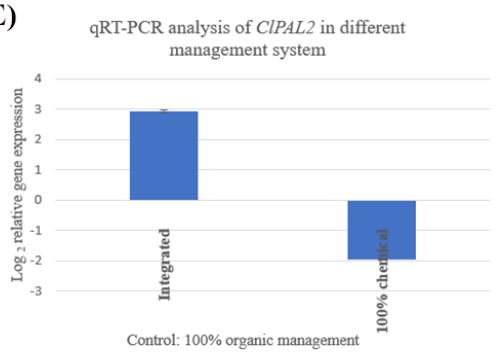
Different locations	Curcumin content(%)
Koz-field	5.37
Erode-field	3.68
Koz-green house	3.01

(D)



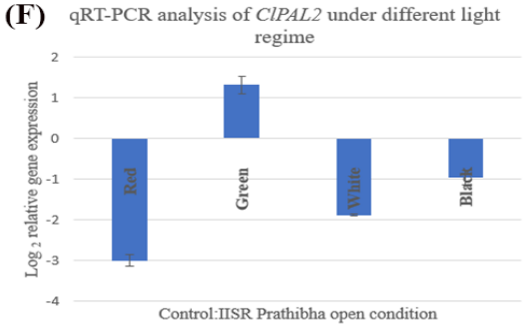
Different accessions	Curcumin content(%)		
	120 DAP	150 DAP	180 DAP
IISR Prathibha	5.37	4.99	3.09
Alleppey supreme	5.21	5.32	4.76
Acc.200	3.54	1.74	3.09
C. aromatica	0.12	0.12	0.48
C.ceasia	0.16	0.04	0.07

(E)



Type of management	Curcumin content (%)
100% organic	4.93
Integrated	5.36
100% chemical	4.23

(F)



Different light regime	Curcumin content (%)
Control (open)	5.03
Red	4.02
Green	4.88
White	4.30
Black	4.59

Figure 31: Relative expression pattern of PAL in (A) high, low and very low curcumin condition for identifiying candidate PAL (B) in different tissues of turmeric (C) different location (D) different developmental stages, (E) different nutrient management, (F) different light regime.

1.4.3.4. Amplification of full length *ClPAL2* cDNA and sequence analysis

Gene cloning is an important tool for not only increasing knowledge of key factors influencing gene regulation but also for integrating structural and functional genomics for improving the knowledge base on pathway regulation. We isolated the 5' and 3' end fragments of the *ClPAL2* cDNA by the rapid amplification of cDNA ends (RACE) technique based on the 961 bp sequence from *Curcuma* transcriptome. A 2500 bp fragment through 5' RACE and a 900 bp fragment through 3' RACE together produced a 2832 bp full-length cDNA of *Curcuma longa* (Figure 32(A) and (B)). Sequence analysis indicated that *ClPAL2* cDNA contained an open reading frame of 2322 bp, a 43 bp 5' terminal untranslated region (UTR), 467-bp in the 3' UTR, and had a poly(A) tail). The ORF encoded a predicted polypeptide of 773 amino acids. Similarly in *Rhus chinensis* ORF was 2,124 bp and encoded a 707-amino acid polypeptide (Ma et al., 2013). Another PAL with open reading frame of 2,142 bp, which encodes a protein of 713 amino acid, and it also contains a 101-bp 59UTR, a 215-bp 39UTR, and a 26-bp polyA tail was identified from Dendrobium using RACE (Jin et al., 2013). Blastp search revealed that the cDNA shared 74.25% identity with *Hedychium coronarium* phenylalanine ammonia lyase (PAL) mRNA, complete cds (QTI57556.1), 68.88% identity to *Musa acuminata subsp. malaccensis* phenylalanine ammonia-lyase-like (XP_009384657.1), 66.80% identity with *Musa acuminata AAA Group* PAL (ACG56648.1). The sequence of *ClPAL2* was deposited in NCBI (MZ818774). The results indicated that the gene identified is novel and closely related to *PAL* from other species of Zingiberaceae. RACE PCR is a good strategy for cloning full length genes especially in cases where genes are long and exist as isoforms and normal cloning strategies are not successful (Jin et al., 2013).

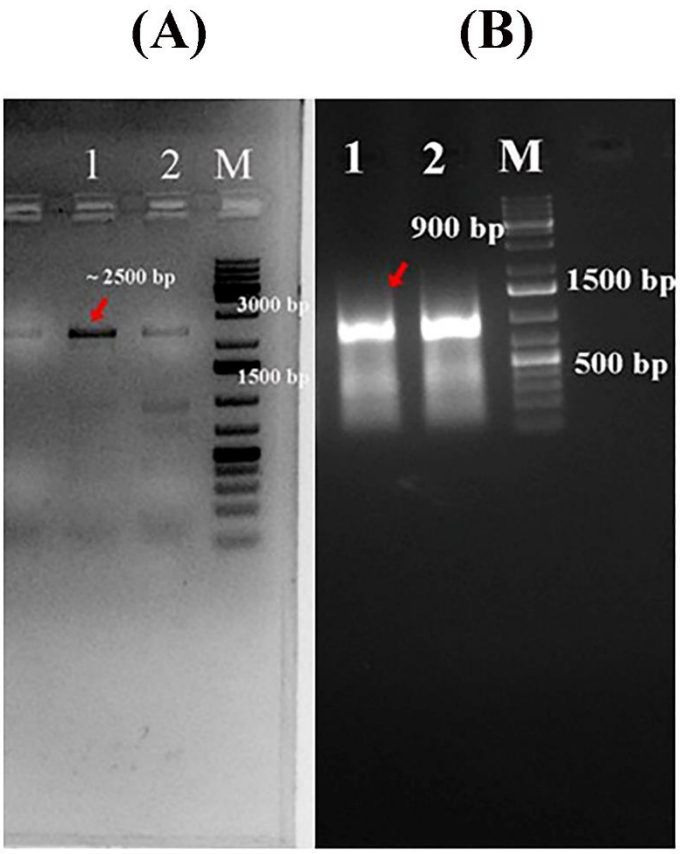


Figure 32: Full-length amplification of PAL (A-5' RACE and B-3' RACE product)

1.4.3.5. Phylogenetic analysis of *ClPAL2*

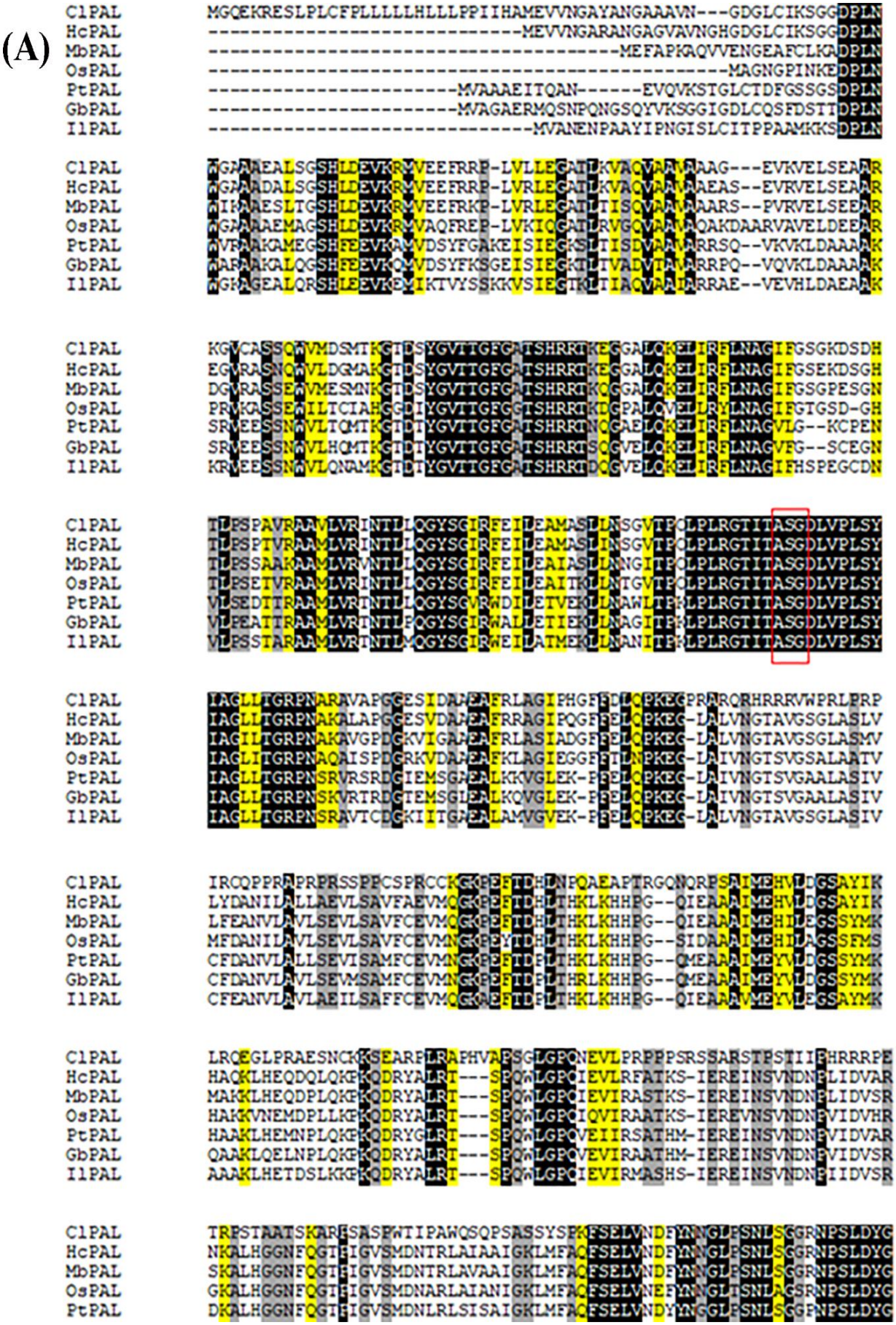
Genomic era has thrown in endless opportunities to annotation and it is prudent to develop phylogenetic relationships that will enable inferring gene functions from those already known, having an identical pattern of expression in phylogenetically diverse genomes. Evolutionary relationship between ClPAL2 and other plant PAL proteins (Table 5) was studied and multiple sequence alignment was done using ClustalW (Thompson et al., 1994) with default parameters (Figure 33 (A)). The conserved amino acid residues from plant PALs were found in ClPAL2 at amino acid positions *viz.*, Y103, L131, A197, S198, G199, N255, Q342, Q342, R345, R349, F395 and Q483 (Figure 33(A)) similar to that reported by Zhu et al., 2015. A phylogenetic tree was constructed by the neighbor-joining method using the MEGA 7.0 program (Figure 33 (B)). ClPAL*2* clustered together with PAL from other species, including GbPAL from *Ginkgo biloba*, PtPAL from *Pinus taeda*, IlPAL from *Isoetes lacustris*, OsPAL from *Oryza sativa japonica group*, MbPAL from *Musa balbisiana*, LrPAL from *Lycoris radiata*, CoPAL

from *Cinnamomum osmophloeum*, VvPAL from *Vitis vinifera*, BnPAL from *Brassica napus*, HcPAL from *Hedychium coronarium* AtPAL, AtPAL2, AtPAL1 from *Arabidopsis thaliana*. In this subgroup, the *ClPAL2* showed maximum identity with *Hedychium coronarium*, a perennial flowering plant of Zingiberaceae. The results indicated that *ClPAL2* is a typical PAL gene derived from *Zingiberaceae* (Ginger) family and may have a similar function as *HcPAL2* of the PPP. This high similarity with functional *PAL* genes from crops of the same family indicates that ClPAL2 is probably involved in the PPP synthesising curcumin. However, the ClPAL2 is clustered together with the other species rather than with a single species. This result indicates that duplication events are important in the evolution of ClPAL2 gene (Kumar and Ellis, 2001)

Table 6: List of sequence used for phylogenetic tree construction

Species	**ID**	**PAL Sequence Names**
Rehmannia glutinosa	AAK84225.1	RgPAL
Solenostemon scutellarioides	JQ975419.1	SsPAL
Salvia miltiorrhiza	EF462460	SmPAL1
Arabidopsis thaliana	AY303128	AtPAL1
Camellia sinensis	D26596.1	CsPAL1
Petroselinum crispum	CAA68938.1	PcPAL1
Jatropha curcas	ABI33979.1	JCPAL
Ricinus communis	XP_002519521.1	RCPAL
Vitis vinifera	XP_002268732.1	VVPAL
Daucus carota	BAC56977.1	DCPAL
Cinnamomum osmophloeum	AFG26322.1	CoPAL
Musa balbisiana	BAG70992.1	MbPAL
Lycoris radiata	ACM61988.1	LrPAL
Arabidopsis thaliana	NP_181241.1	AtPAL
Brassica napus	ABC69916.1	BnPAL
Ginkgo biloba	ABU49842.1	GbPAL
Oryza sativa Japonica Group	CAA34226.1	OsPAL
Medicago sativa	CAA41169.1	MsPAL
Solanum lycopersicum	AAA34179.2	SlPAL
Nicotiana tabacum	BAA22948.1	NtPAL
Petroselinum crispum	CAA57056.1	PcPAL
Arabidopsis thaliana	NP_181241.1	AtPAL
Nicotiana tabacum	BAA22948.1	NtPAL
Pinus taeda	AAA84889.1	PtPAL
Isoetes lacustris	AAW80637.1	IlPAL
Prunus avium	AAC78457.1	PaPAL
Hedychium coronarium	QTI57556.1	HcPAL

(A)



(B)

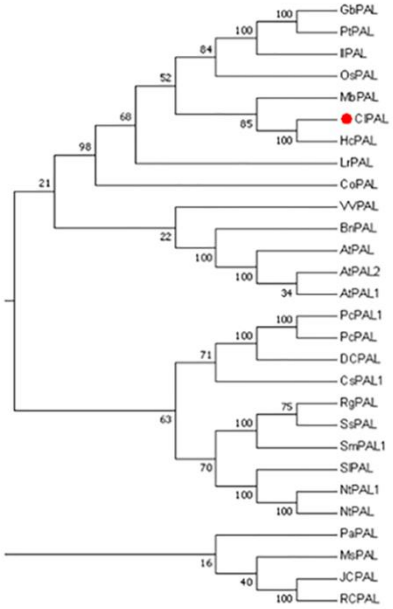


Figure 33: Comparison of ClPAL2 with other plant PALs by (A) Multiple sequence alignment and (B) Phylogenetic analysis

1.4.3.6. The structure of ClPAL2 protein and sequence analysis

The predicted amino acid sequence of ClPAL2 was highly similar to the reported plant PAL protein sequence. ClPAL2 protein was further compared with homologous proteins by means of multiple sequence alignment and structure prediction analysis. The result showed that ClPAL2 contains the conserved PAL protein finger motif (228-243 GTITASGDLVPLSYIA), which is a typical PAL/HAL protein tag and the catalytic triad residues (Ala 232, Ser 233 and Gly234). By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction revealed that α-helices (45.54%) were the main structural elements, and random coils (37.13%) were scattered in the entire protein (Figure 34 (A)). The active site of PAL was comparable to that of histidine ammonia lyase (HAL), whose structure was determined using X-ray crystallography (Ritter and Schulz, 2004). Structure of PAL consisting of the MIO domain, a core domain and an inserted shielding domain like a “sea-horse” shape (Mahesh et al., 2006, Ritter and Schulz, 2004) (34(B). In the MIO domain, there was a highly-conserved Ala-Ser-Gly triad, which acts as the active site forming the MIO prosthetic group for non-oxidative deamination (Lister et al., 1996). By using the known *Petroselinum crispum* PAL (PcPAL; 6f6t.1.A) (Ritter and Schulz, 2004) as a template in the SWISS-MODEL

program (Arnold et al., 2006), the tertiary structure of ClPAL2 (Figure 35 (A)) was deduced and the model was validated through Ramachandran plot (Figure 35(B)). These data imply that *ClPAL2* might be a functional protein with catalytic activity similar to PcPAL.

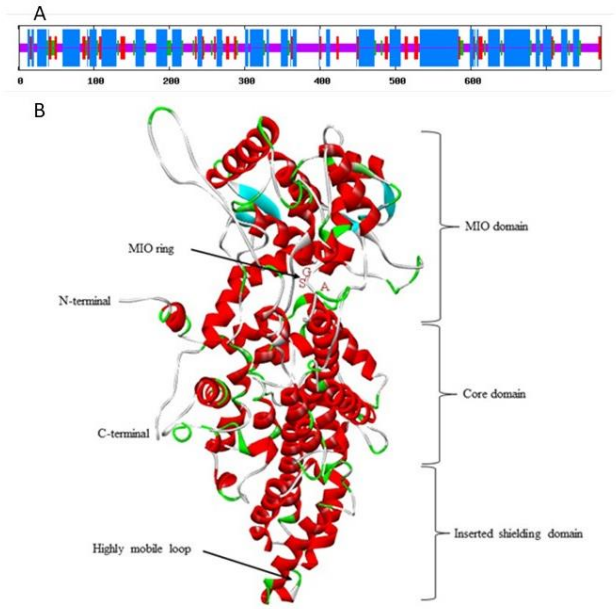


Figure 34: Predicted molecular model of *ClPAL2*. (A) Predicted secondary structure of *ClPAL2*. The red, green, blue and pink regions represent, extended strand, beta turn, alpha helix and random coil, respectively. (B) Predicted tertiary structure of *ClPAL2*. The α-helices, β-sheets, turns and coils are shown in red, cyan, dark and white, respectively. The MIO ring, Ala-Ser-Gly, is marked red.

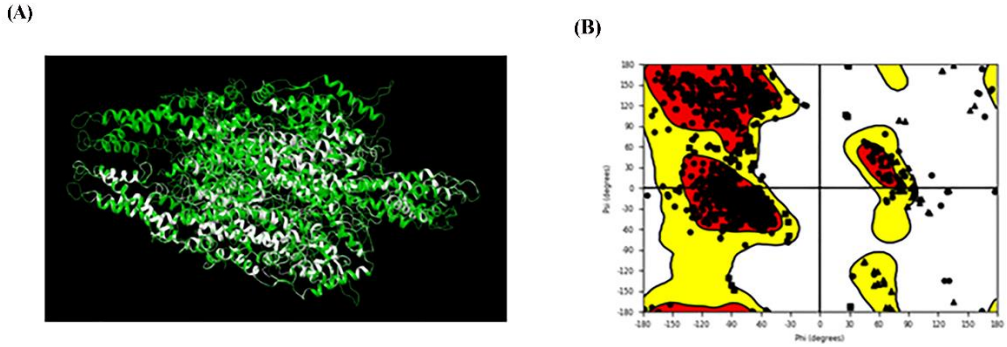


Figure 35: Molecular modelling of *C. longa ClPAL2* (A) 3-D structure and (B) Ramachandran plot determined by PROCHECK.

1.4.3.7. Molecular docking of *ClPAL2*

The tool predicted 5 binding pockets, and selected the binding pocket based on the highest C-score value of 0.87, (The confidence score of the prediction C-score ranges [0-1], where higher score indicates a more reliable prediction. There are 11 active residues are present on the predicted sites, they are 138,144,15,166,233,236,290,416,441,521 and 525. These amino acids are formed as the cavity for ligand binding. Grid was generated with AutoDock 4.2 and the target proteins were embedded in the grid. XYZ coordinates were adjusted while setting grid for enclosing the active residues. It ensures the accurate binding of proteins with ligand molecules. The grid spacing was given as 0.375 and the xyz dimension as 60 ×60× 60 for fourteen proteins. The grid was centred on the active site and XYZ-coordinates of the macromolecules were 124.926 A^0, 18.438 A^0 and 22.956 $A^{0.}$ All other parameters were kept as default for docking. Docking results showed that substrates docked satisfactorily to the enzyme active site of ClPAL2 with good docking scores (Figure 36) and lower energy scores represent better protein-ligand bindings compared with higher energy values (Thomsen and Christensen, 2006) (Table 7).

Table 7: Docking result of *ClPAL2* with phenylalanine

Binding energy [ΔGbind (kcal/mol)]	-3.87	
Inhibitor constant [KI (µm)]	1.45	
H-bond	THR142	ARG289
Bond type	NH…O	OH…O
Bond length (A^0)	2.15	1.89

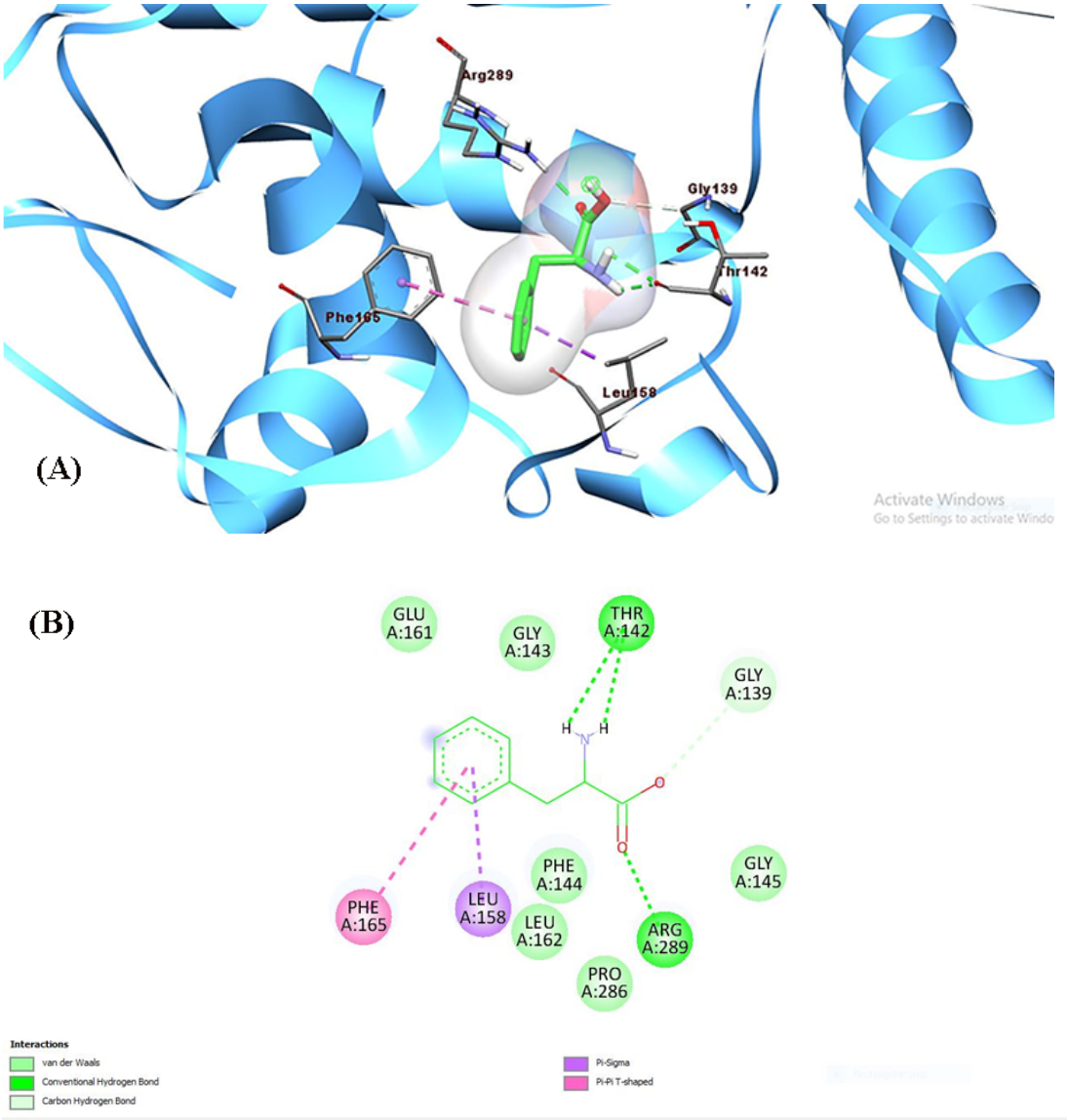


Figure 36: (A) Docking interaction of phenylalanine ammonia lyase in the active site of ClPAL2 and (B) 2D diagram of the phenylalanine-ClPAL2 interaction.

1.4.3.8. Identification of candidate genes and isoforms of O-methyl transferases

The chemical structure of curcumin contains two O-methyl groups (Priyadarsini, 2014b). According to Kita et al., 2008 O-methylation takes place after the formation of the curcuminoid skeleton (Kita et al., 2008). However, it is not known when the methyl ethers at the 3' position of curcumin are incorporated, *i.e.*, before or after curcumin scaffold formation (Katsuyama et al., 2009). The transcriptome data was highly instrumental in identifying the key O-methyltransferase gene transcripts involved in curcumin biosynthesis. Thirty O-methyltransferases were present in the transcriptome of PC and PS that showed differential expression w.r.t curcumin. Among them three were caffeoyl CoA O-methyltransferases and rest belonged to caffeic acid OMT (COMT)

subfamily. Different types of FOMTs with different levels of substrate specificity and regioselectivity are cooperatively involved in the flavonoid biosynthesis (Itoh et al., 2016). It could be possible that several O-methyltransferase unigenes may be responsible for the production of these diverse phenylpropanoid compounds (Bureau et al., 2007). Methylation improves the solubility and activities of phenylpropanoid compounds (Koirala et al., 2016). Several caffeic acid O-methyltransferases were identified earlier in phenylpropanoid biosynthesis pathway in many crops like tobacco (Maury et al., 1999), ice plant involved in phenylpropanoid biosynthesis pathway was identified from *Vitis vinifera* (Giordano et al., 2016) and *Citrus reticulata* (Liu et al., 2020b). Among the thirty related transcripts five that showed high FPKM in PC compared to PS was selected for further analysis (Table 8). The five contigs were c41020_g1_i2 (FPKM in PC: 37.67; PS: 1.23), c41020_g1_i1 (FPKM in PC: 25.04; PS: 0.57), c41020_g1_i3 (FPKM in PC: 21.18; PS: 0.85), c42730_g1_i2 (FPKM in PC: 108.08; PS: 3.55), and c32328_g1_i1 (FPKM in PC: 12.4; PS: 9.88). Contig c42730_g1_i2 showed the maximum FPKM in PC (108.08) compared to PS (3.55) and c32328_g1_i1 showed minimum expression in PC compared to PS. c42730_g1_i2 showed 67.29% similarity with already reported *trans-resveratrol di-O-methyltransferase-like [Elaeis guineensis]* (XM_008790173.3).

All those five isoforms of OMT identified by us were up regulated in IISR Prathibha (≥5.5% curcumin), grown in field conditions and showed >65% similarity with the already reported plant O-methyltransferases. The transcript c41020_g1_i2 showed 99.18%% similarity with coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*] (KJ780360.1). Similarly, two isoforms namely c41020_g1_i1, c41020_g1_i3 showed similarity with already reported coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*] (KJ780360.1) with percentage similarities of 98.92% and 98.51 respectively. (Table 8). Caffeoyl-CoA O-methyltransferase involved in flavonoid biosynthesis were identified through comparative transcriptome analysis in *Ampelopsis megalophylla* (Yang et al., 2019). Most of the studies reported in candidate selection strategy of O-methyltransferase were based on amino acid sequence identity compared with functionally characterized O-methyltransferases (Chang et al., 2015). Very limited reports of comparative transcriptome analysis for candidate selection of O-methyltransferases are available (Foong et al., 2020; Yang et al., 2019).

Table: 8 Gene ontology analysis of O-methyltransferase genes showing differential expression in PC compared to PS

Contig ID	Sequence length	FPKM value in PC	FPKM value in PS	Sequence description	Blast top hit description
c41020_g1_i2	1124	37.67	1.23	coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*]	KJ780360.1
c41020_g1_i1	1205	25.04	0.57	coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*]	KJ780360.1
c41020_g1_i3	406	21.18	0.85	coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*]	KJ780360.1
c42730_g1_i2	1121	108.08	3.55	PREDICTED: trans-resveratrol di-O-methyltransferase-like [*Elaeis guineensis*]	XM_008790173.3
c32328_g1_i1	366	0.17	9.88	PREDICTED: glucuronoxylan 4-O-methyltransferase 1-like [*Musa acuminata subsp. malaccensis*]	XM_009417622.2

1.4.3.9. Expression pattern of *OMT* vs curcumin content

The expression patterns of five OMT isoforms were studied in four-month-old rhizomes of IISR Prathibha (5.37% curcumin), Acc 200 (3.50% curcumin) and *C. aromatica* (0.12% curcumin). ClPKS11 were used as the bait gene (Deepa et al., 2017; Deepa 2018) and *C. aromatica* was kept as control. Among those c42730_g1_i2 (fold change-20.67) and c41020_g1_i2 (fold change-75.58) showed enhanced expression in high curcumin accession and low in low curcumin accessions (fold change-1.03 and 2.29 in c42730_g1_i2 and c41020_g1_i2 respectively), exactly correlating with the pattern of expression of the bait gene, ClPKS11(fold change-40 and 0.12 in high and low curcumin accessions respectively). Other contigs such as c41020_g1_i1 (fold change-2 and 0.48 in high and low curcumin accessions respectively), c41020_g1_i3 (fold change-6.54, 3.8 in low and high curcumin accessions respectively) and c32328_g1_i1 (fold change-2.73 and 1.08 in high and low curcumin accession respectively) also showed a similar pattern,

though the expression was comparatively less. (Figure 33 (A)). This suggests that these contigs might be members of multigene families or part of same gene without overlapping sequences and might integrate into single large contigs (Upadhyay et al., 2014).

Contig c42730_g1_i2, a caffeic acid OMT and c41020_g1_i2, a caffeoyl CoA OMT were selected for further validation as they showed maximum expression in correlation with curcumin. The expression patterns were evaluated in different tissues such as leaves, pseudo-stem, root, and rhizomes under field conditions at Kozhikode at 120 DAP with root as control. Both *ClOMT2* and *ClOMT3* were upregulated in rhizome (fold change-13.26 and 831.74 respectively) followed by leaf (fold change-1.05 and 661.08 respectively) and pseudostem (fold change-0.53, 37.27 respectively). Results showed that ClOMT2 expression correlated with curcumin whereas ClOMT3 expressed uniformly in all tissues at high expression levels irrespective of curcumin status. This may be due to the fact that caffeoyl CoA O-methyltransferases are involved in upstream biosynthesis pathways that are common to phenylpropanoids compounds in general (Le Roy et al., 2016), whereas caffeic acid O-methyltransferase is specific to the final downstream steps immediately leading to synthesis of curcumin (Figure 37 (A)). Tissue specific analysis of caffeoyl CoA O-methyltransferases reported in *Vanila planifolia* also indicated a tissue specific differential expression (Widiez et al., 2011). Tissue specific expression analysis in *S cultivar* the key genes involved in the anthocyanin biosynthesis were significantly upregulated in specific tissue (petals) of the, which was consistent with the high accumulation of anthocyanins in *S. cultivar* petals.

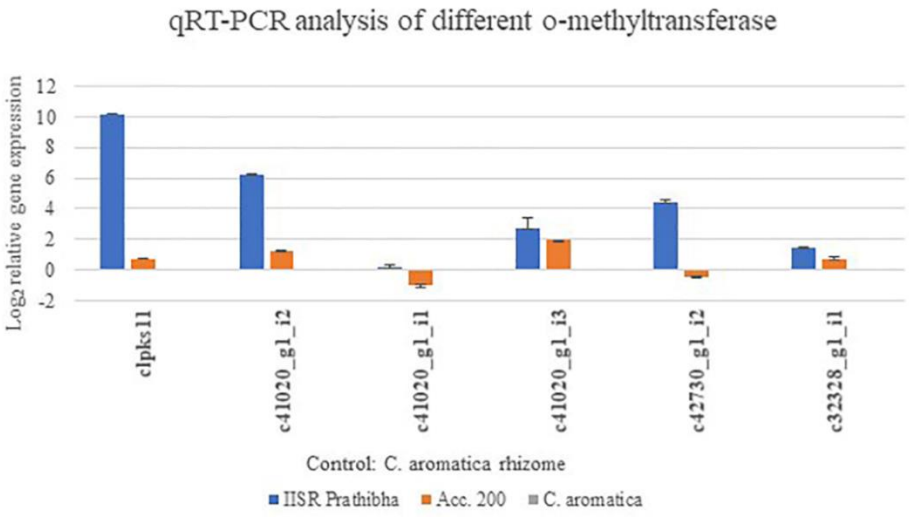
Evaluation under different geographical locations indicated a maximum curcumin content under field conditions at Kozhikode (Koz-field (≥5.5%), followed by Erode field (3.68%) and Koz-green house (3.01%). Maximum expression of ClOMT2 was observed in Koz-field followed by Erode-field (fold change-0.11) and least expression was observed in Koz-green house (fold change-0.01). This indicated a positive correlation of gene expression with curcumin in ClOMT2. Expression of ClOMT3 was maximum at Koz-green house (fold change-0.60) compared to Erode-field (fold change-0.08), which did not correlate with curcumin status (Figure 37(C)). Results thus indicate that curcumin biosynthesis is influenced by environmental variations (Sasikumar, 2005; Anandaraj et al., 2014). Our results suggest that the agro-climatic condition in Kozhikode, where turmeric is rain fed favoured the production of curcumin in IISR Prathibha.

Environmental factors were reported to influence anthocyanin biosynthesis in many plants (Feng et al., 2010; Li et al., 2012). In apples and pears, low temperature favours anthocyanin accumulation and expression of genes related to anthocyanin biosynthesis (Ubi et al., 2006; Steyn et al., 2005).

Temporal expression analysis of OMTs under different developmental stages (120 DAP, 150 DAP, and 180 DAP) in IISR Prathibha (>5.5% in 120 DAP, 4.99% in 150 DAP and 3.09% in 180 DAP), IISR Alleppey Supreme (5.21% in 120 DAP, 5.32% in 150 DAP and 4.76% in 180 DAP), Acc. 200 (3.54% in 120 DAP, 1.74% in 150 DAP and 3.09% in 180 DAP), *C. aromatica* (0.12% in 120 DAP, 0.12% in 150 DAP and 0.48% in 180 DAP) and *C. ceasia* (0.16% in 120 DAP, 0.04% in 150 DAP and 0.07% in 180 DAP) with Acc 200 as a control, indicated highest expression of ClOMT2 (fold change-18.50) and ClOMT3 (fold change-8.11) were found in IISR Alleppey supreme followed by IISR Prathibha (fold change-10.33) at 120 DAP. Expression of ClOMT2 (fold change-15.45) slightly decreased at 150 DAP with further reduction at 180 DAP (fold change-6.68) whereas expression of ClOMT3 was drastically reduced at 150 DAP (fold change-1.98) and again reduced at 180 DAP (fold change-1.12). Expression of ClOMT2 (fold change-0.18 and 0.11 in *C. aromatica* and *C. ceasia* 120 DAP) and ClOMT3 was down-regulated in both *C. aromatica* (0.12% curcumin) and *C. ceasia* (0.16% curcumin) at all developmental stages, which correlates with the low curcumin in these related species (Figure 37(D)). This indicates that the accumulation of these enzymes is correlated with curcumin biosynthesis. This is similar to studies by Sun et al., 2015; Huang et al., 2015; Shi et al., 2014 in which the gene expression profile was correlated to metabolite accumulation. Similar study related to anthocyanin was done in Roses indicate that several classic flower pigmentation-related genes were highly correlated with their corresponding metabolites (Lu et al., 2021). Similarly, another study conducted in *Acer mandshuricum* which involved combined transcriptome and metabolome integrated analysis to reveal candidate genes involved in anthocyanin accumulation result indicated that metabolite correlated well with transcript. Thus, the expression of ClOMT2 and ClOMT3 correlated with curcumin content in all the experiments in the present study, which indicates that the novel gene identified could be involved in regulation of curcumin biosynthesis in turmeric. It has been reported that curcumin content varies in different varieties (Sasikumar, 2005; Sajitha et al., 2014) and subject to agro-climatic variations (Anandaraj et al., 2014; Singh et al., 2013). Metabolic

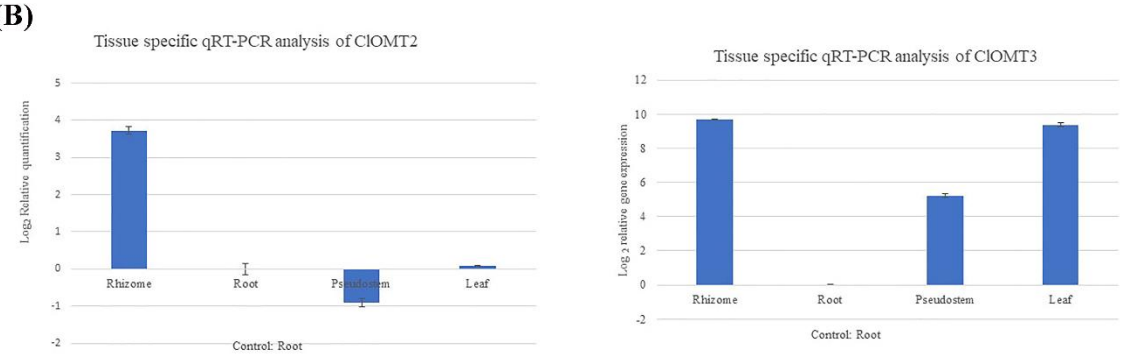
changes are affected mainly through a single or a few enzymes, whose activity is most likely to lead to notable changes in metabolite concentrations. The activity of an enzyme may be related to various factors, including environmental conditions that may cause a change in gene expression. Therefore, the analysis of expression levels, metabolite concentration, and their connections has attracted considerable attention (Zhang et al., 2021).

(A)

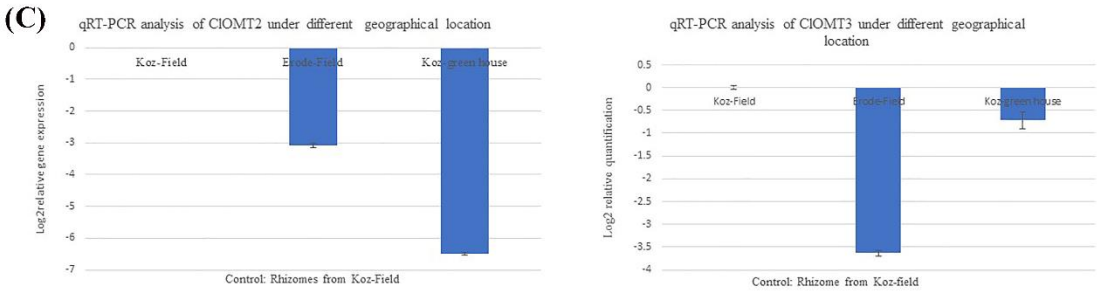


Different turmeric accessions	Curcumin content(%)
IISR Prathibha	5.37
Acc. 200.	3.50
C. aromatica	0.12

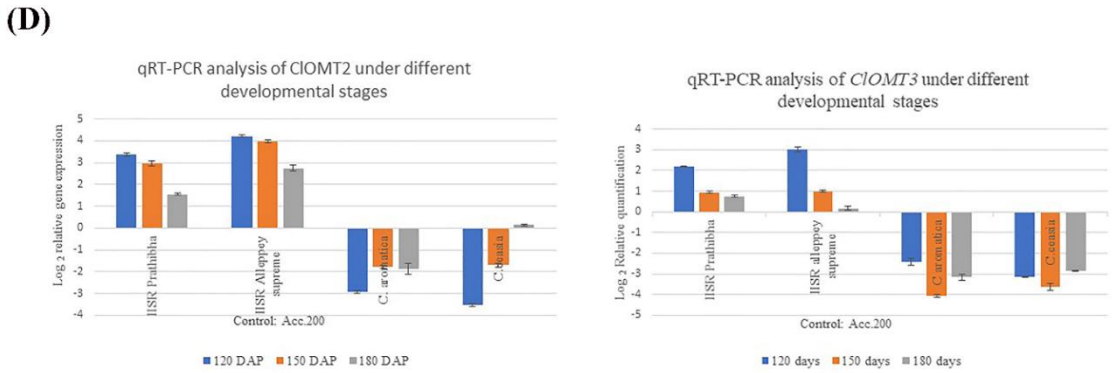
(B)



Different locations	Curcumin content(%)
IISR Prathibha rhizome	5.37



Different locations	Curcumin content(%)
Koz-field	5.37
Erode-field	3.68
Koz-green house	3.01



Different accessions	Curcumin content(%)		
	120 DAP	150 DAP	180 DAP
IISR Prathibha	5.37	4.99	3.09
Alleppey supreme	5.21	5.32	4.76
Acc.200	3.54	1.74	3.09
C. aromatica	0.12	0.12	0.48
C.ceasia	0.16	0.04	0.07

Figure 37: Relative expression pattern of OMT in (A) high, low and very low curcumin condition for identifiying candidate OMT (B) in different tissues of turmeric (C) different location (D) different developmental stages

1.4.3.10. Amplification of full length *ClOMT2* and *ClOMT3* cDNA and sequence analysis

Complete ORF of *ClOMT2* and *ClOMT3* was confirmed by cloning based on primers designed against the contig id c4730_g1_i2 and c41020_g1_i2 respectively.The PCR product of size ~ 1100 bp and 1500 bp was obtained from genomic DNA as the template and 1000 bp and 750 bp was obtained with cDNA as template. These genes were designated as *ClOMT2* and *ClOMT3* (*C. longa* OMT). *ClOMT2* contains two exons connected by an intron and *ClOMT3* contain five exons connected by an intron

(Figure 38). The sequence of *ClOMT2* (OK631847) and *ClOMT3* (OK631848) were deposited in NCBI. Similarly 1,373 bp long and encodes a 386-amino-acid protein

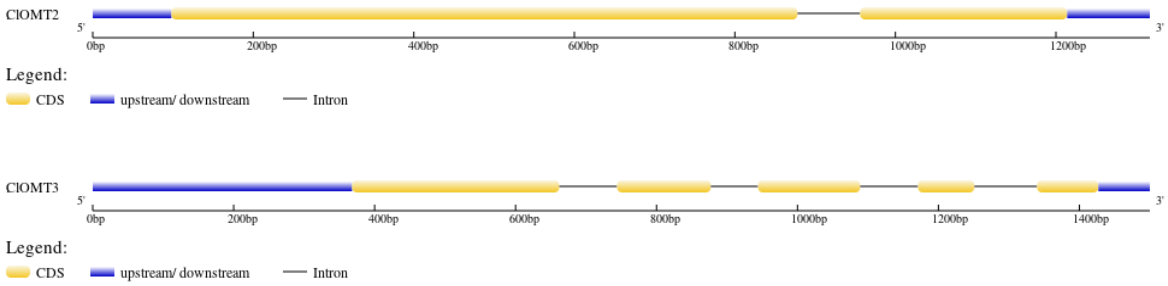


Figure 38: Gene distribution of ClOMT2 and ClOMT3

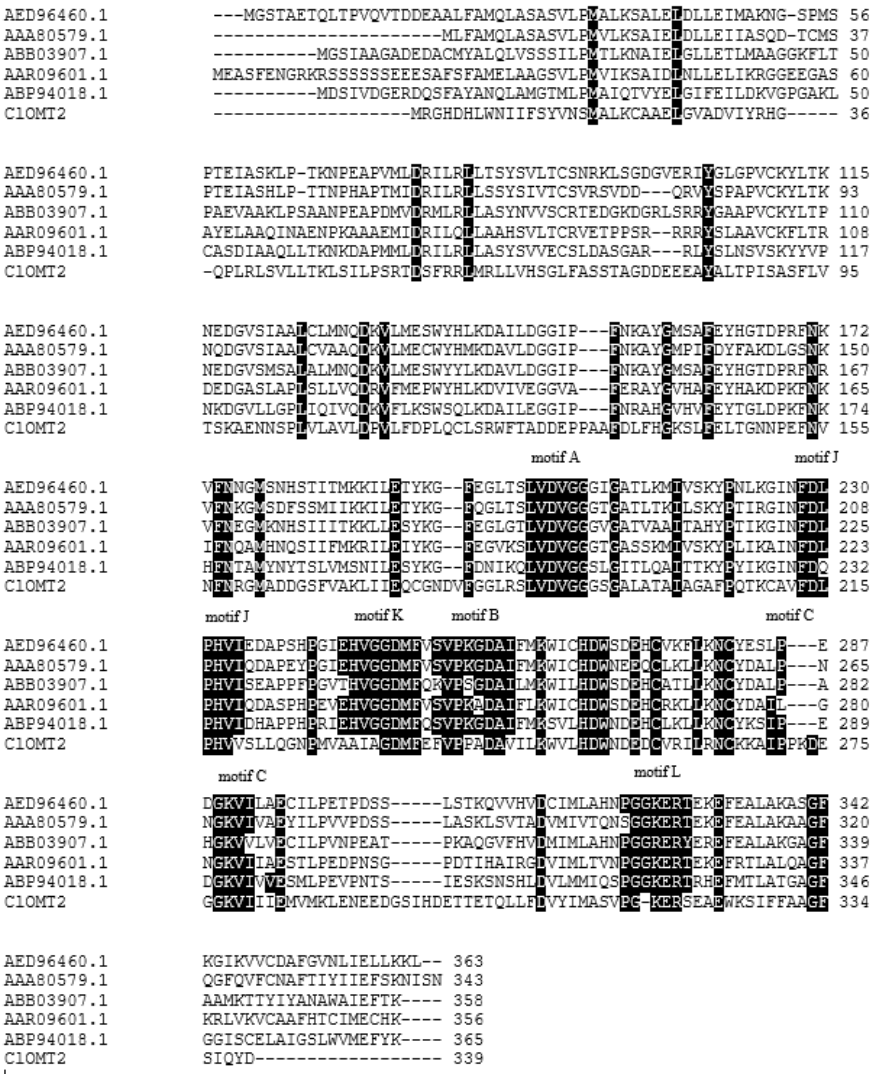


Figure 39: Multiple sequence alignment of ClOMT2. Highly conserved regions were marked in black. Different motifs such as motif A, motif J, motif K, motif B, motif C and motif L were also marked

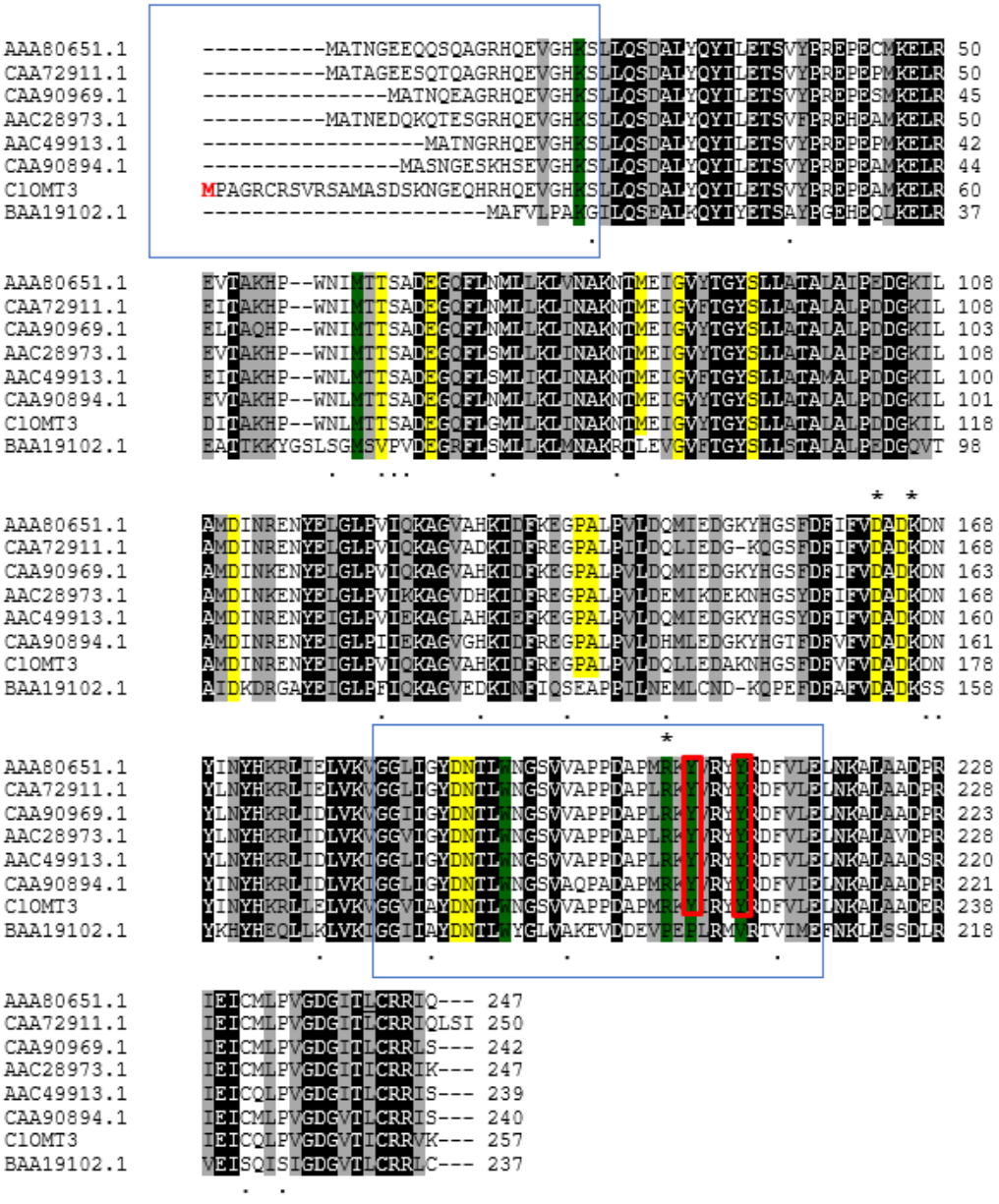


Figure 40: Multiple sequence alignment of ClOMT3. Highly conserved regions were marked in black. Residues involved in divalent metal ion and cofactor binding are depicted on a yellow background. Residues involved in substrate recognition are depicted on a green background. An asterisk indicates identical residues in all sequences, ash colour indicates strongly conserved residues (score>0.5), and a ‘‘.’’ indicates weaker conserved residues (score<0.5)

1.4.3.11. Phylogenetic analysis of *ClOMT2* and *ClOMT3*

To study the evolutionary relationship of *ClOMT2* and *ClOMT3* and other OMT proteins, a set of OMT amino acid sequences from other plants were used for multiple sequence alignment using ClustalW (Thompson et al., 1994) employing default parameters and different motifs specific to the O-methyltransferase and caffeoyl CoA were identified (Figure 39 and 40). Phylogenetic tree was constructed by the neighbor-joining method using the MEGA 7.0 program (Figure 41). *ClOMT2* closely cluster with O-methyltransferase from *Citrus depressa,* flavonoid O-methyltransferase from *Triticum aestivum,* O-methyltransferase from *Arabidopsis thaliana,* 3'-flavonoid O-methyltransferase *Chrysosplenium Americanum,* flavonoid 3'-O-methyltransferase from *Mentha piperita* whereas *ClOMT3* clustered together with caffeoyl-CoA O-methyltransferases from *Pinus taeda, Petroselinum crispum, Nicotiana tabacum, Medicago sativa,* and *Vitis vinifera.* This indicates the diversity in this gene available in turmeric.

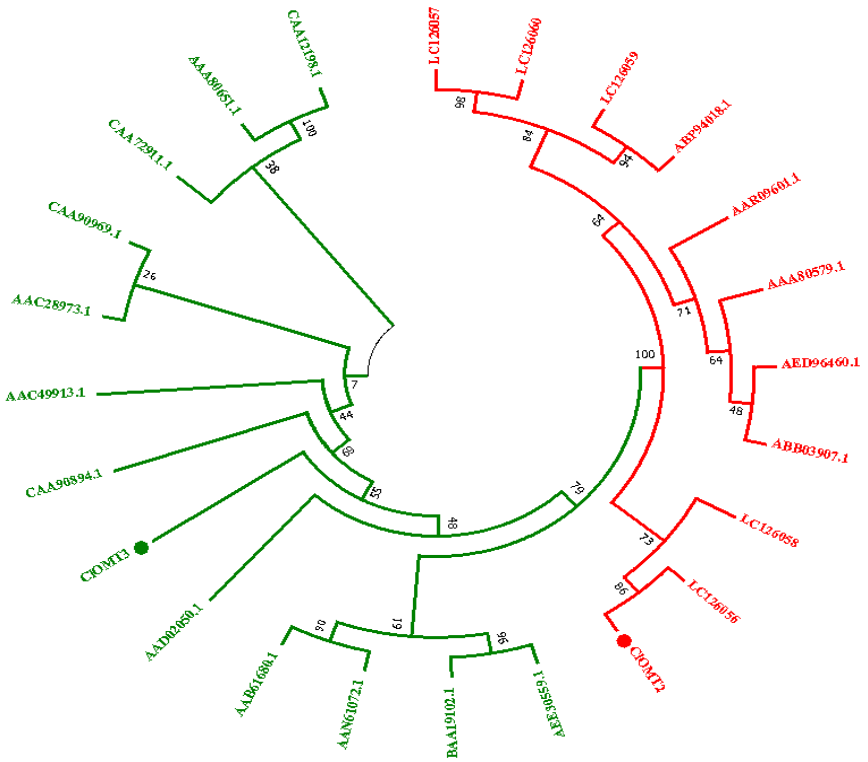


Figure 41: Phylogenetic analysis of ClOMT2 and ClOMT3 genes from *Curcuma longa* by MEGA 7.0. The two groups were distinguished in different colors and the numbers on the branches were Bootstrap values.

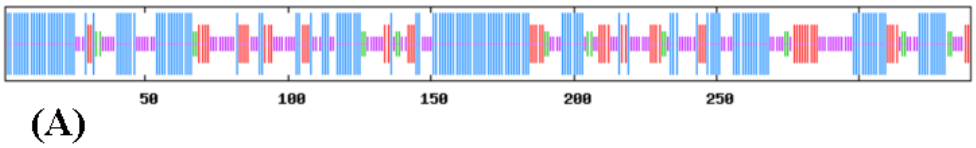
1.4.3.12. The structure of ClOMT2 and ClOMT3 protein, Sequence analysis and molecular docking

The predicted amino acid sequence of ClOMT2 was 90% similar to trans-resveratrol di-O-methyltransferase-like from *Zingiber officinale*. The *ClOMT2* protein was further compared with homologous proteins by means of multiple sequence alignment and structure prediction analysis. *ClOMT2* contains the different conserved motifs such as motif A, motif J, motif K, motif B, motif C and motif L (Figure 39). By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction of *ClOMT2* revealed that α-helices (46.31%) were the main structural elements, and random coils (31.86%), extended strands (15.93%) and β turns (5.90%) were scattered in the entire protein (Figure 42 (A). S)-norcoclaurine 6-O-methyltransferase (5icc.1. A) with 38.49% identity was used as a template in the SWISS-MODEL program. The 3D structure of ClOMT2 protein model (Figure 42 (B)) was confirmed using Ramachandran Plot statistics (Figure 42 (C)). The predicted amino acid sequence of ClOMT3 was highly similar (99.18%) to the reported coffee acyl coenzyme A-3-O-methyl transferase from *Curcuma longa*. The ClOMT3 protein was further compared with homologous proteins by means of multiple sequence alignment and structure prediction analysis. The result showed that ClOMT3 contains the conserved residues involved in divalent metal ion and cofactor binding and residues involved in substrate recognition (Figure 43). By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction of *ClOMT3* revealed that α-helices (39.18%) were the main structural elements, and random coils (35.92%) and β turns (7.35%) were scattered in the entire protein (Figure 43 (A). By using the known caffeoyl-CoA O-methyltransferase from *sorghum* (*CCoAOMT*; *5kva.1. A*) with 84.49% identity as a template in the SWISS-MODEL program. The 3D structure of ClOMT3 protein model (Figure 43 (B)) was confirmed using Ramachandran Plot statistics (Figure 43 (C)).

For ClOMT2 the COACH-D server predicted 5 binding pockets, and choose the binding pocket based on the highest C-score value of 0.87, (The confidence score of the prediction C-score ranges [0-1], where higher score indicates a more reliable prediction. There are 17 active residues are present on the predicted sites, they are 143,156,160,164,190,191,213,214,217,233,234,235,247,248,249,252 and 253. These

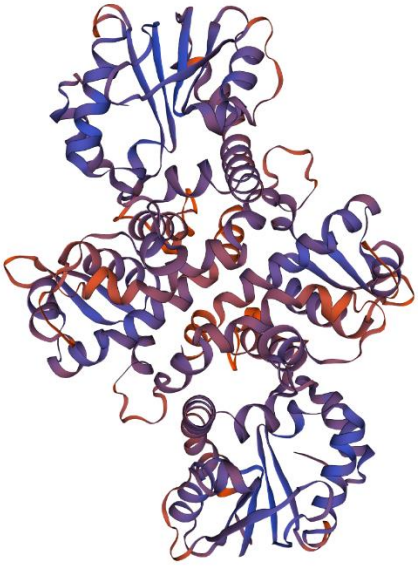
amino acids are formed as the cavity for ligand binding. The grid was generated with AutoDock 4.2 and the target proteins were embedded in the grid. XYZ coordinates were adjusted while setting grid for enclosing the active residues. It ensures the accurate binding of proteins with ligand molecules. The grid spacing was given as 0.375 and the xyz dimension as 60 ×60× 60 for fourteen proteins. The grid was centred on the active site and XYZ-coordinates of the macromolecules were -6.879 A0, 3.133 A0 and 26.851 A0. All other parameters were kept as default for docking. To check whether single O-methyltransferase is responsible for the methylation of final conversion of curcumin. Docking analysis of ClOMT2 performed with two substrates. And bisdemethoxycurcumin and demethoxycurcumin docked with the protein molecule to study which ligand molecule having the least binding energy. Docking results showed that the demethoxycurcumin almost same binding energy. And lowest binding energy score indicating that better protein-ligand bindings and substrates docked satisfactorily to the enzyme active site of ClOMT2 with good docking scores (Figure 44 and 45 and table 9).

For ClOMT3 the COACH-D server predicted 5 binding pockets, and choose the binding pocket based on the highest C-score value of 0.99, (The confidence score of the prediction C-score ranges [0-1], where higher score indicates a more reliable prediction. There are 23 active residues are present on the predicted sites, they are 41,42,43,67,68,69,72,73,90,91,92,93,96,118,119,120,121,141,143,144,145,148 and 152. These amino acids are formed as the cavity for ligand binding. The grid was generated with AutoDock 4.2 and the target proteins were embedded in the grid. XYZ coordinates were adjusted while setting grid for enclosing the active residues. It ensures the accurate binding of proteins with ligand molecules. The grid spacing was given as 0.375 and the xyz dimension as 60 ×60× 60 for fourteen proteins. The grid was centred on the active site and XYZ-coordinates of the macromolecules were 16.067 A0, 47.726 A0 and 6.873 A0. All other parameters were kept as default for docking. Docking results showed that substrates docked satisfactorily to the enzyme active site of ClPAL3 with good docking scores (Figure 46, Table 9) and lower energy scores represent better protein-ligand bindings compared with higher energy values. Molecular docking is the simple method for identification of substrate (Favia et al., 2007).



(A)

(B)



(C)

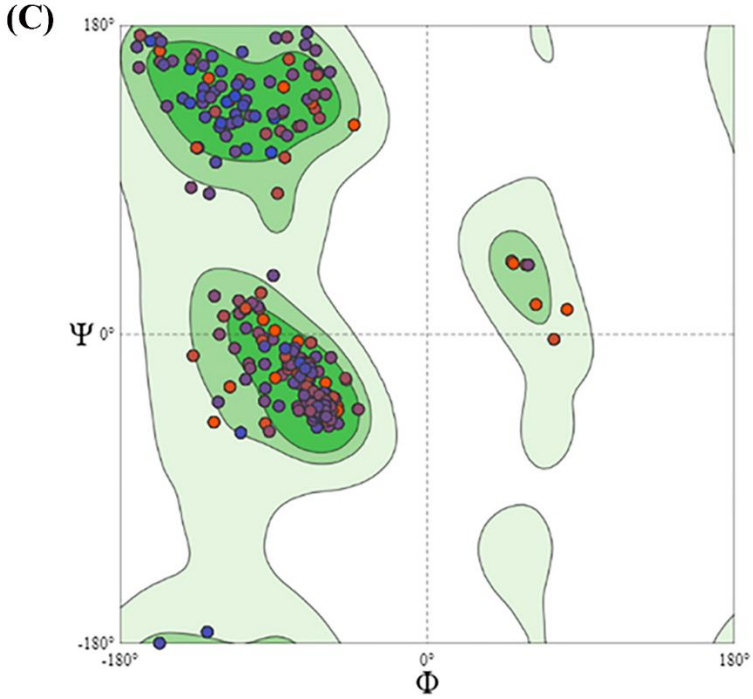
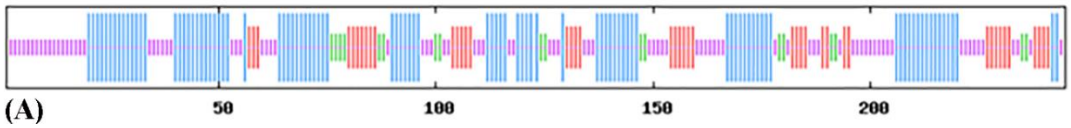
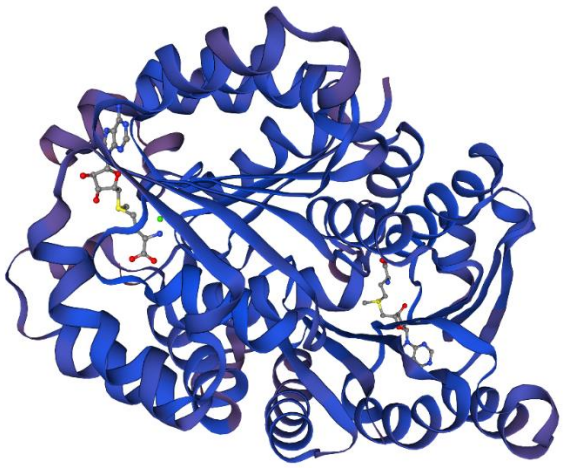


Figure 42: (A) Predicted secondary structure of ClOMT2, (B) molecular modelling of *C. longa* CLOMT2 and 3-D structure representation and (C) Ramachandran plot determined by PROCHECK.



(B)



(c)

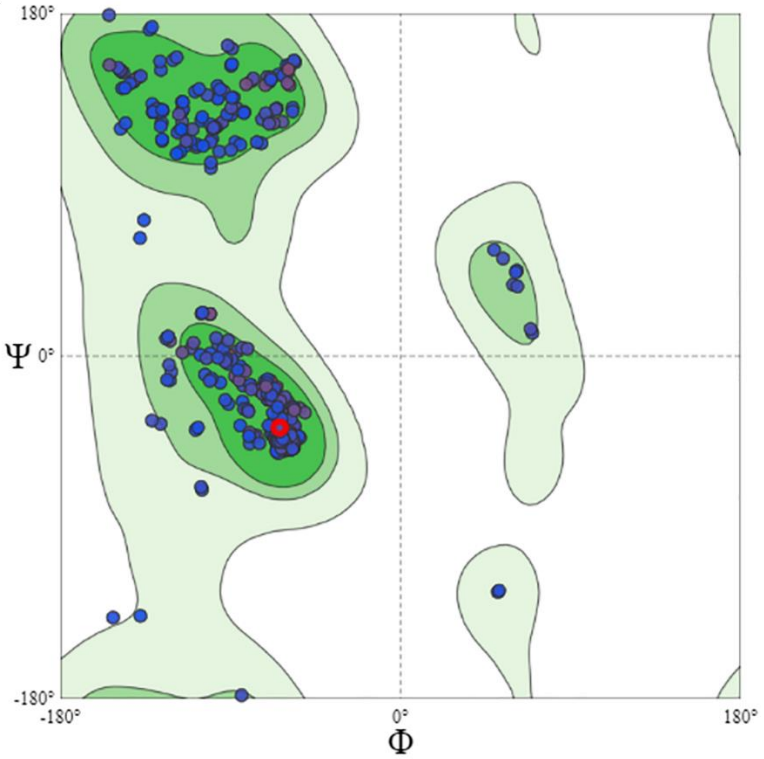


Figure 43: (A) Predicted secondary structure of ClOMT3, (B) molecular modelling of *C. longa CLOMT3* and 3-D structure representation (C) Ramachandran plot determined by PROCHECK.

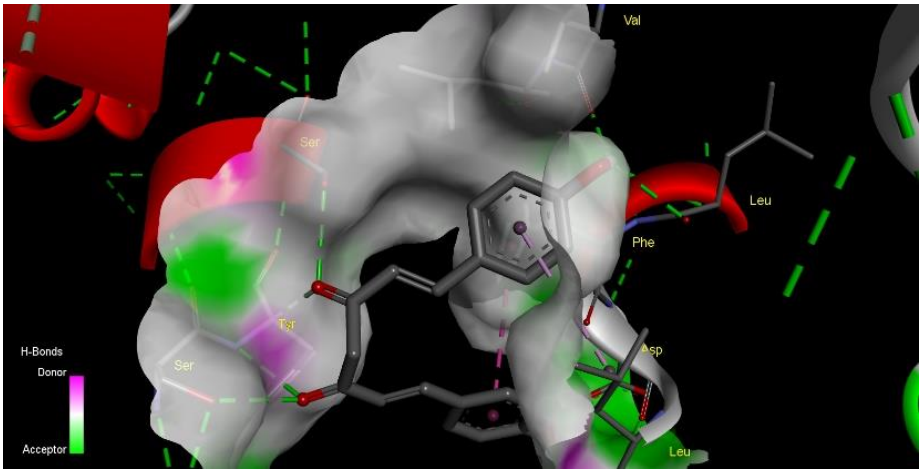


Figure 44: Docking interaction of ClOMT2 with bis-dimethoxy curcumin

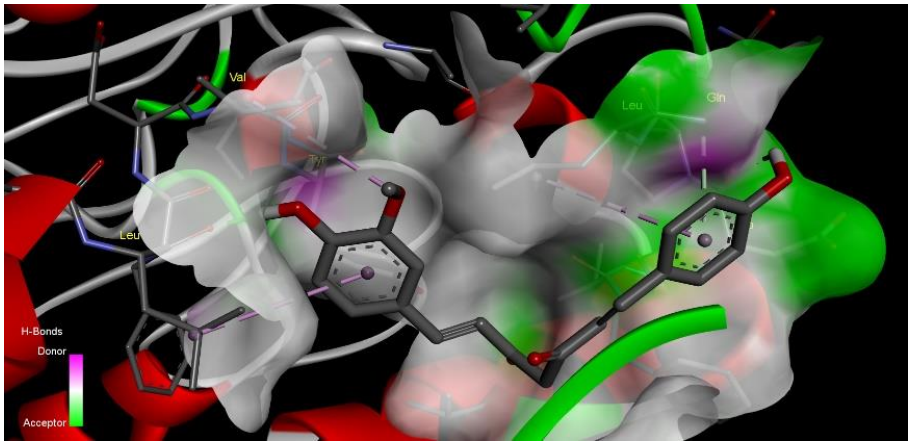


Figure 45: Docking interaction of ClOMT2 with demethoxy curcumin

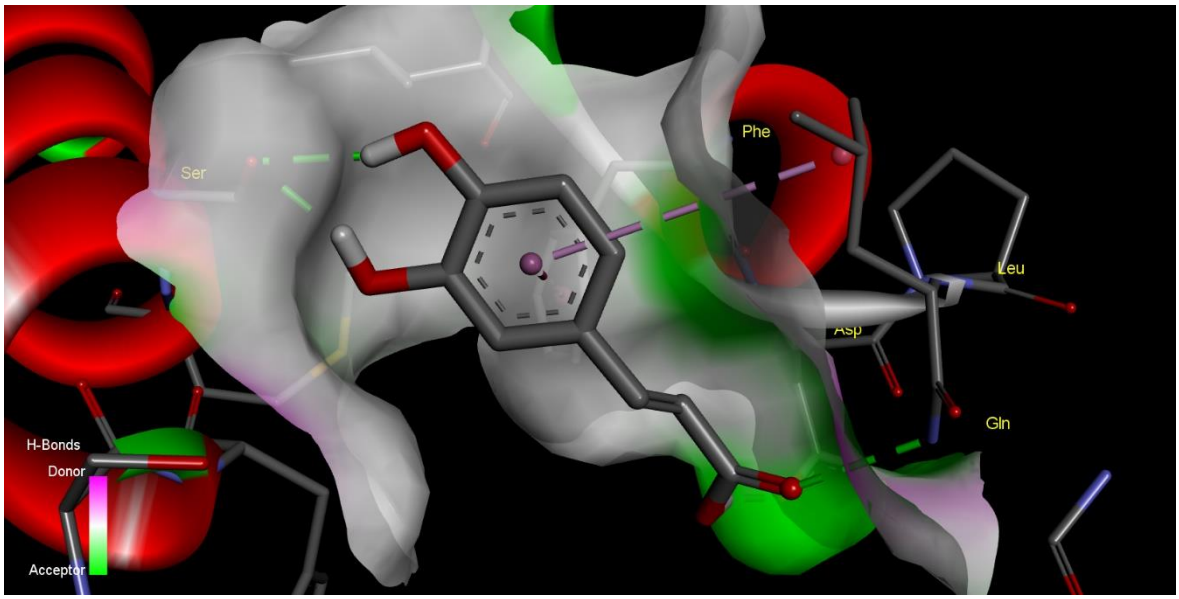


Figure 46: Docking interaction of ClOMT3 with caffeic acid

Table 9: Docking result of ClOMT2 and ClOMT3 with its substrate

Protein molecules	Ligand molecules	ΔGbind (kcal/mol)	KI (μm)	H-bond	Bond type
Caffeoyl CoA O-methyltransferase	Caffeoyl CoA	-1.27	117.59	Cys230 ASP236 ASN192 THR60 CYS230	H:O H:O HN:O H:O H:O
O-methyltransferase B chain	demethoxycurcumin	-5.98	41.56	GLN119	H:N
O-methyltransferase B chain	bismethoxycurcumin	-5.83	53.16	ASP116 SER13 VAL109	H:O OG:O H:O
O-metyltransferase _caffeic acid	Caffeic acid	-3.68	1.43	SER17 SER17 GLN119 SER17	H:OG OG:O N:O H:OG

1.5 CONCLUSION

Transcriptome sequencing could be judiciously used as a cost-effective method for developing genomic information in non-model species. RNA-seq data generated in our study enable identification of DEGs involved in specific pathway of curcumin, paralogs of genes involved in the pathway as well as identified molecular biomarkers for curcumin. A higher percentage similarity of transcriptome was observed with related member of Zingiberaceae. The functional annotation indicated maximum genes grouped under ATP binding, translation and integral component of membrane allowing identification of novel genes involved in curcumin biosynthesis. A large number of PPP genes directly related to curcumin showed differential expression in high and low curcumin lines and correlated well with curcumin levels. These genes were shortlisted by us as associated to curcumin biosynthesis. This, study has focused on only two important genes *PAL* and *OMT* due to limitations in time and resources. The candidate isoform

involved in the pathway were identified by co-expression analysis that correlated gene expression to curcumin content under different field experiments. Invariably all the identified candidate genes correlated with curcumin under all conditions. These genes were cloned fully, structure of gene and protein analysed by *in silico*, and docking studies, to further support the co-expression analyses. The study enabled optimization of several techniques like SMARTER PCR, RACE PCR, co expression models *etc.* which may be extended to other candidate genes and crops.

Publication/ abstract

- **Prashina Mol P,** Aparna R S, Deepa K and Sheeja T E (2019). "A novel Phenylalanine ammonia lyase gene identified from *Curcuma longa* L. could be a bonafide candidate involved in biosynthetic regulation of curcumin" organized by Central Coffee Research Institute on 6th to 8th March 2019, pp-8.
- **Prashina Mol P**, Aparna RS, Sheeja TE, Deepa K. "A novel o-methyltransferase gene from *Curcuma longa* L.: A putative structural gene regulating curcumin biosynthesis" SYMSAC-X during 09-12 February 2021, ICAR-Indian Institute of Spices Research. (Received best Research Paper Award (Poster)-Dr. Alapatti Prasad Rao Award).

Chapter 2

TF AND miRNA MEDIATED REGULATION OF CURCUMIN BIOSYNTHESIS

2.1. INTRODUCTION AND OBJECTIVE

The main structural genes involved in the biosynthetic pathway for curcuminoids are phenylalanine ammonia lyase (*PAL*), cinnamate - 4 - hydroxylase (*C4H*), 4-coumaryl CoA ligase (*4CL*), diketide CoA synthase (*DCS*) and curcumin synthases 1, 2 & 3 (*CURS1*, *CURS2*, *CURS3*) (Katsuyama et al., 2009a; 2009b). Apart from those, certain novel structural genes like *CLPKS11*(Deepa et al., 2017, *ClPAL2*, *ClOMT 2* and *ClOMT3* (Chapter 2) were identified and their role in biosynthesis established through co expression studies in field trials (Deepa et al., 2017; Deepa 2018; Mol et al., 2021; Aparna et al., 2021). The expression of the structural genes is regulated at the transcriptional level by TFs. The identification of promoter elements by transcription factors is essential for getting an insight into transcriptional regulation of gene expression. TFs interact with the regulatory regions of the target genes and control transcriptional initiation process through up regulation or down regulation (Vom Endt et al., 2002). In some cases, they interact with other TFs to form a complex to regulate gene expression (Yang et al., 2012, Zhao et al., 2019). Several families of TFs including bHLH, WD40, MYB, WRKY, bZIP, NAC *etc*. have been identified to regulate plant secondary metabolism in various crops (Zhao et al., 2013; Saga et al., 2012; Xu et al., 2004; Wang et al., 2015). Biosynthesis of phenylpropanoids is generally controlled by a gene regulatory network involving MYB transcription factors regulating early biosynthetic steps and a complex of MYB-bHLH-WD40 TFs regulating the late biosynthetic steps (Li et al., 2014). Late biosynthetic steps are more specific to the final metabolite and are more important for pathway engineering. We have already reported MYB TFs regulating curcumin biosynthesis in an earlier study (Sheeja et al., 2015); however, no information on key TFs such as bHLH and WD40 are available. Hence, in the present study, we evaluated bHLH and WD40 TFs and identified three novel TFs that negatively regulate curcumin biosynthesis, which may lead to unravelling the regulatory mechanism.

MicroRNAs (miRNAs) are short noncoding RNAs (22–24 nucleotides) that regulate gene expression by partnering with matched target mRNA and promoting

mRNA degradation and/or translation inhibition (Llave et al., 2002). Several investigations have shown that miRNA can control the biosynthesis of phenylpropanoid (Gupta et al., 2017; Marcela et al., 2019; Sharma et al., 2016). The majority of miRNAs have been found to target structural genes and transcription factors and negatively regulate at post transcriptional level. For the discovery of new and species-specific miRNA, computational strategies have proven effective (Zhang et al., 2006). We had identified conserved, non-conserved and novel miRNAs from turmeric through *in silico* prediction and Illumina deep sequencing (Santhi and Sheeja, 2015). To uncover the role of miRNA in curcumin biosynthesis we used Illumina based deep sequencing of two accessions contrasting in curcumin content. The studies in our lab have indicated that curcumin biosynthetic pathway is highly regulated and the structural genes possess DNA binding sites at upstream regions with binding sites for transcription factors like MYBs. Our earlier transcriptome studies and differential expression analysis identified several TFs which showed a positive correlation with the expression of candidate pathway genes (Sheeja et al., 2015; Deepa et al., 2017). We could also identify a number of novel miRNAs which regulated MYB TFs of PPP (Santhi et al., 2018; Sheeja et al., 2018). All these inspired us to explore these regulatory molecules for their probable role in controlling expression of pathway genes. Hence a study was formulated with the objective to identify and characterize regulatory molecules of curcumin biosynthesis pathway.

2.2. REVIEW OF LITERATURE

Transcription factors (TFs) are proteins that regulate gene expression by specifically binding to cis-acting promoter region of the target gene. Transcription factors have DNA-binding domains (DBDs) that identify specific DNA sequences in promoters. However, there are regulatory proteins that do not have a DBD but interact directly with DNA-binding TFs to form transcriptional complexes and are also classified as transcription factors (Hong, 2016). Based on the DNA binding domains TFs were classified into different families (Riechmann et al., 2000). AP2-ERF, NAC, Dof, YABBY, WRKY, GARP, TCP, SBP, ABI3-VP1 (B3), EIL, and LFY are plant-specific TF families. Major transcription factors that are conserved across the eukaryotes are bZIP TF family, bHLH TF family, MYB TF family and HSF TF family. Plant specific transcription factor family includes AP2/ERF family, WRKY TF family, NAC TF family, and TCP TF family. B-box zinc finger TFs, AUX/IAA family, JAZ/protein

family are the transcription factors without DNA binding domains however they interact with DBD containing transcription factors (Hong, 2016). Basic region/leucine zipper (bZIP) TFs have motif for protein dimerization and a basic region for DNA binding. In *Arabidopsis* bZIP TFs regulate diverse processes like defence responses, light signalling, stress responses and flower development (Jakoby et al., 2002). bZip transcription factor have role in phenylpropanoid biosynthesis pathway such as anthocyanin and flavonoid biosynthesis pathway. A bZIP transcription factor which activates flavonoid biosynthesis pathway was identified from *Vitis vinifera* L (Malacarne et al., 2012). Another bZIP transcription factor named MdHY5 identified from apple which bind to E box motif of its own promoter and G box motif of MdMYB10 and positively regulated the both expression and thereby regulating anthocyanin biosynthesis (An et al., 2017). Basic helix-loop-helix (bHLH) TFs are the largest and widely distributed transcription factor family among eukaryotes. bHLH family transcription factors have two functionally distinctive region named basal region and HLH region. Basic region contains approximately 15 amino acids and it located at the N terminal of bHLH domain function as DNA binding domain. HLH region help to form homo/hetro dimers and which contains two aliphatic α group with linking loop of variable length(Atchley et al., 1999). Co-expression of GtMYB3 and GtbHLH1 are involved in the accumulation of gentiodelphin, an anthocyanin in gentian flowers which responsible for the blue colour of the flower were identified (Nakatsuka et al., 2008).Three bHLH transcription factors (ZjGL3a, ZjGL3b, and ZjTT8) were found to be correlated with anthocyanin content in Jujube Fruit during fruit development (*Ziziphus Jujuba* Mill.) (Shi et al., 2019). The MYB family are ubiquitous and present in functionally diverse forms in all eukaryotes. Which contains a conserved DNA binding domain known as MYB domain help to bind DNA. The domain consists of total of 52 amino acids contain four imperfect amino acid sequence repeats (R), each forming a helix-turn-helix structure that interact with the DNA major groove. Depending on the number of MYB repeats (one to four) MYB proteins are divided into different classes (Dubos et al., 2010; Katiyar et al., 2012; Riechmann et al., 2000). Three major groups pf plant MYB TFs are available namely R2R3 MYB (two adjacent repeat), R1R2R3 MYB (three adjacent repeat) and MYB related proteins. Several R2R3 MYB transcription factors were involved in phenylpropanoid biosynthesis (Li et al., 2017, 2019; Liu et al., 2016; Ma et al., 2018). Several MYB transcription factors were involved in regulation of phenylpropanoid biosynthesis in both monocotyledons and dicotyledons. MYB transcription factors

regulate anthocyanin biosynthesis genes in *Zea mays* (Mehrtens et al., 2005), *Oryza sativa* (Zheng et al., 2019), *Arabidopsis thaliana* (Zhou et al., 2015), *M. domestica* (Chagné et al., 2013). TFs like WRKY belong to a large class in plants with a typical 60 amino acid long, four stranded β-sheet WRKY DNA binding domain along with a zinc finger motif. Based on the characteristics of the DNA binding domain WRKY are divided in to three groups. Group I with two WRKY DBD, group II with single DBD with different C2H2 zinc finger and group III consisting of single DBD with C2HC zinc finger (Rushton et al., 2010). A Novel WRKY Transcription Factor (HmoWRKY40) was identified through Yeast one-hybrid and transient expression assays and was found to bind and activate the promoter of HmoCYP76AD1 gene responsible for the betalain biosynthesis in Pitaya (*Hylocereus monacanthus*) (Zhang et al., 2021). A WRKY transcription factor (AaWRKY40) involved in terpenoid biosynthesis was identified in *Artemisia annua* (Paolis et al., 2020). NAC transcription factors are plant specific transcription factors involved in plant development, biotic and abiotic stress. (Ooka et al., 2003). The name NAC was derived from consensus sequence from petunia (No apical meristem (NAM)), *Arabidopsis* (ATAF1/2), and cup-shaped cotyledon2 (CUC2) proteins. NAC domain is 150 amino acid long and contains five conserved regions termed as A-E (Hong, 2016). A novel NAC transcription factor was identified from apple (*Malus domestica* Borkh. and over-expression of MdNAC42 in apple calli resulted in the up-regulation of flavonoid pathway genes thereby increasing the accumulation of anthocyanins (Zhang et al., 2020). The WD40 proteins (WD repeats) are heteromeric with several copies of WD repeats with each repeat containing 40-60 residues. Each unit contains 11-24 residues of glycine histidine dipeptide at N-terminal and Trp-Asp (WD) and double residue at the C-terminus (Neer et al., 1994; Smith et al., 1999).

A protein complex composed of MYB and bHLH transcription factors associated with a WD40 repeat protein initiates multiple cellular differentiation pathways in a range of plants. The MYB-bHLH-WD40 (MBW) complex is formed by the interaction of R2R3-MYB-TFs with bHLH and the WD40 repeat (WDR) (Li et al., 2019). The MYB protein is thought to be responsible for the particular activation of target genes. In a variety of plant species, the transcription complex WD40, bHLH, and MYB modulates the expression of numerous different target genes. Flavanol biosynthesis is controlled by three R2R3-MYB proteins (MYB11, MYB12, and MYB111), whereas anthocyanin and PA biosynthesis is controlled by the MYBbHLH-WD40 (MBW) complex, which

activates late biosynthetic genes (Li, 2014). R2R3-MYBs usually interact with bHLH1 and WD40 to create an MBW activation complex, thereby increasing the expression of bHLH2 and the structural genes to promote the accumulation of anthocyanin (Albert et al., 2011; Moglia et al., 2020) (Figure 1).

In tomato (*S. lycopersicum*)) anthocyanin pigments were enhanced in fruits when MBW complex activates the expression of the SlAN1(bHLH) gene together with most of the anthocyanin structural genes (Yan et al., 2020). The MBW complex produced by the interaction of PsMYB12 with bHLH and WD40 activated PsCHS, a key gene involved in tree peony (*Paeonia suffruticosa*) pigmentation, whereas the lack of CHS activity results in albino flowers (Durbin et al., 1995, 2000; Gu et al., 2019).

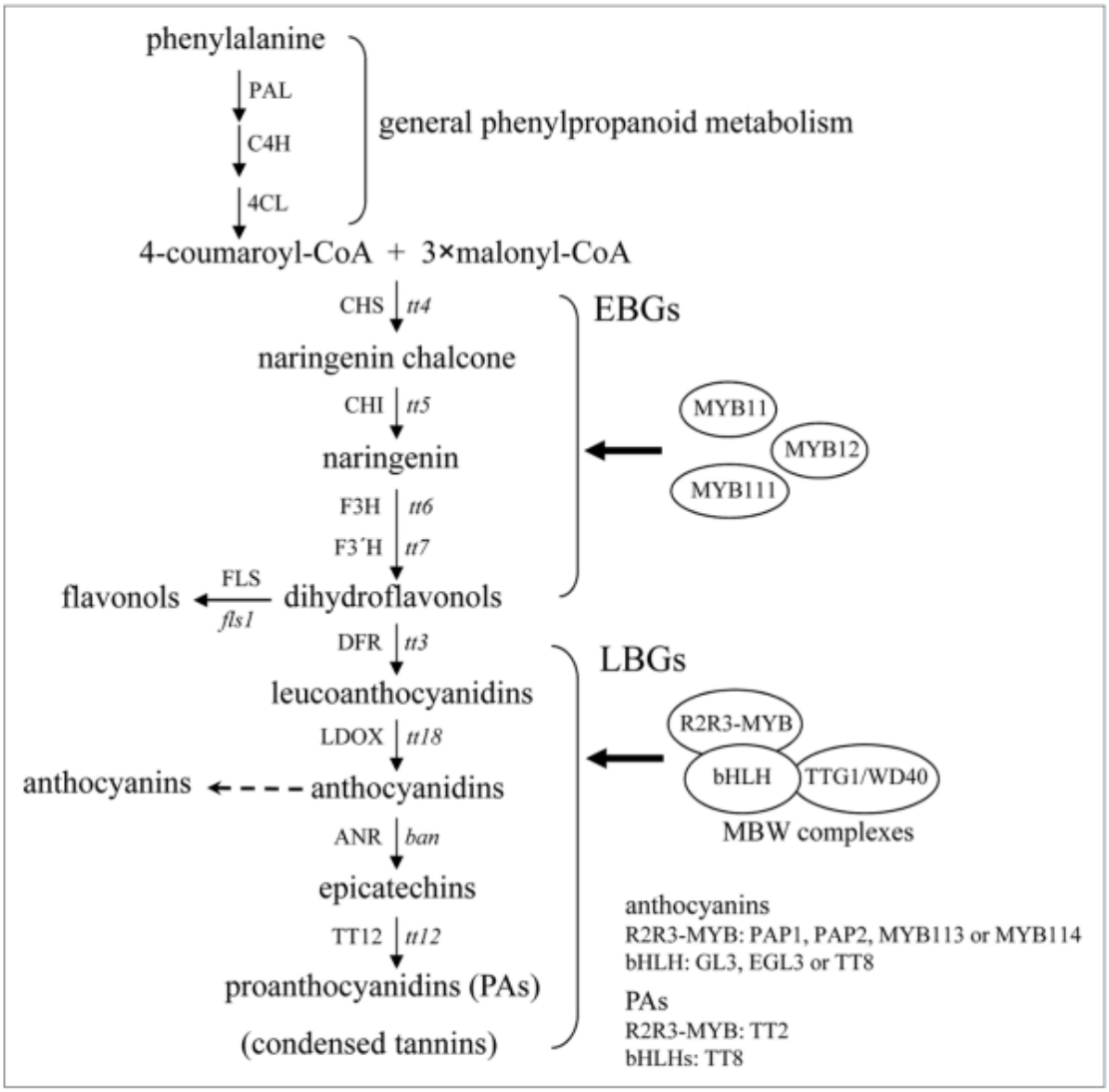


Figure 1: The biosynthetic pathway for flavonols, anthocyanins, and Pas (Proanthocyanins) in Arabidopsis (Li, 2014).

Lin-4 was the first miRNA discovered in *Caenorhabditis elegans*, however 22-nt lin-4 and 21-nt let-7 were discovered later in the same organism (Melo and Melo, 2014). MicroRNAs (miRNAs) are found in the genome as single or clustered polycistronic units that share function relationships (Bruscella et al., 2017). Most miRNAs are encoded by their own primary transcript in plants, especially from introns which act as the hotspots for their origination. Target of miRNA could be 37 miRNAs are identified from turmeric by *in silico* method (Manila, 2009). miRNA can regulate the different targets from same family or a single member of a gene family. miRNAs regulate the expression of genes at the post-transcriptional level *via* mRNA degradation and/or translational repression (Catalanotto et al., 2016). In plants, there is a perfect complementarity between miRNA and target mRNA (Catalanotto et al., 2016) and plant miRNAs control the expression of genes that encode transcription factors, stress response proteins, and others, which have an impact on biological processes. The miRNAs regulate the important biological processes such as primary and secondary metabolism (Tchatchou et al., 2017). Currently, a few studies have revealed the phenomenon of miRNA mediated regulation of secondary metabolic pathways (Gupta et al., 2017). Computational analysis carried out in two transcriptomes of *Swertia* resulted in the identification of miRNAs associated with secondary metabolite biosynthesis; miR-156a, miR-166a, miR-166b, miR-168, miR-11071, and miR-11320 targeting metabolic enzymes, such as aspartate aminotransferase, ribulose-phosphate 3-epimerase, acetyl-CoA acetyltransferase, phosphoglycerate mutase, and premnaspirodiene oxygenase, also include a gene encoding a homeobox-leucine zipper protein (HD-ZIP) with a possible association in secondary metabolite biosynthesis in *Swertia chirayita* (Padhan et al., 2016). Accumulation of anthocyanins in the stems of *A. thaliana* was found to be regulated by a miR156-targeted squamosa promoter binding protein-like (SPL) genes. High miR156 activity promotes accumulation of anthocyanins and activity-induced by flavonols. This study also demonstrates that SPL9 negatively regulates anthocyanin accumulation through destabilization of a MYB-bHLH-WD40 transcriptional activation complex. *Diospyros kaki* fruits collected at two stages (15 and 20 WAF) showed differential expression of mRNAs, indicating that these miRNAs might regulate PA synthesis during development, and some of them are miR858 and miR156, which regulate PA synthesis. miR858 positively regulates the genes responsible for the production of PA, while miR156 downregulates them. miR395 is another miRNA that has an influence on PA biosynthesis (Luo et al., 2015). Some miRNAs (U436803, U977315, U805963,

U3938865, and U4351355) regulate fatty acid and flavonoid biosynthesis in *Lonicera japonica* (Liu et al., 2017; Liu et al., 2017a). The characterization in *A. thaliana* shows that miR858a targets MYB transcription factors that are involved in flavonoid biosynthesis, growth, and development. Over-expression of miR858a down-regulates several MYB transcription factors, and the higher expression of MYBs in MIM858 lines leads to the redirection of the metabolic flux toward the synthesis of flavonoids (Sharma et al., 2016). Yang et al., 2015 indicate that miRNAs regulate some salt stress-related biological pathways includes calcium signaling pathway, MAPK signaling pathway, plant hormone signal transduction, and flavonoid biosynthesis (Yang et al., 2015). miR1438 target caffeoyl-CoA O-methyltransferase related to phenylalanine metabolism in Himalayan mayapple (*Podophyllum hexandrum*), while miR1438 target caffeoyl-CoA O-methyltransferase related to phenylalanine metabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis, stilbenoid, diarylheptanoid, and gingerol biosynthesis. miR1873 targets dihydroflavonol 4-reductase C related to flavonoid biosynthesis, while miR5532 regulates 2-hydroxyisoflavanone dehydratase related to biosynthesis of isoflavonoid (Biswas et al., 2016). miR-156a, miR-166a, miR-166b, miR-168, miR-11071, and miR-11320 targets metabolic enzymes, such as aspartate aminotransferase, ribulose-phosphate 3-epimerase, acetyl-CoA acetyltransferase, phosphoglycerate mutase, and premnaspirodiene oxygenase, also include a gene encoding a homeobox-leucine zipper protein (HD-ZIP) with a possible association in secondary metabolites biosynthesis in *Swertia chirayita*.

2.3. MATERIALS AND METHODS

2.3.1. Functional characterisation of Transcription factors

2.3.1.1. Plant material

Rhizomes were collected after 120 days of planting from high-curcumin turmeric- IISR Prathibha, low curcumin turmeric- Accession No. 200 and a closely related species practically devoid of curcumin- *C. aromatica* from the Experimental Farm, Chelavoor of ICAR-Indian Institute of Spices Research (ICAR-IISR), Kozhikode, Kerala State, India (Samples mentioned in Chapter 2, Section 2.4.1.). To study the spatio-temporal, tissue specific expression and nutrient management expression of TFs, samples and procedures mentioned in Chapter 2 section 2.4.1 and 2.4.2 were used.

2.3.1.2. Screening of *Curcuma* transcriptome for different transcription factor unigenes

Homologous transcripts of different types of TFs like bHLH, WD40, NAC, WRKY, bZIP with fold change >3 was retrieved from the text of annotated genes within transcriptome data of *C. longa* (high curcumin) and *C. aromatica* (practically devoid of curcumin) (Sheeja et al., 2015) (Table 1). Gene ontology (GO) terms were retrieved for each unigene using the default settings of Blast 2GO software (Conesa and Gotz, 2008).

2.3.1.3. cDNA synthesis and quantitative real-time PCR

Total RNA was isolated as per Deepa et al. (2014), and three biological replicates were pooled for each sample and was digested using an on-column DNase I digestion set. The quality and quantity of RNA were checked by agarose gel electrophoresis (1.2%) and Biophotometer plus (Eppendorf, Germany). The RNA samples showed two discrete bands of 28S and 18S rRNA on agarose gel and A260/280 ratio between 1.8 and 2.0. About 250 ng of total RNA was used for first-strand cDNA synthesis using Superscript III reverse transcriptase (Invitrogen) and Oligo-(dT)18 primer in a total volume of 20 µL. Gene expression profiling was performed on Rotor-Gene Q (Qiagen) using the QuantiFast SYBR Green PCR kit (Qiagen). The reaction mixture comprised of 10 µL of 2X SYBR Green, 10 pmol each of gene-specific primers and 100 ng of cDNA in a final volume of 20 µL. qRT-PCR amplification was performed under the following conditions: 95^{o}C for 5 min, followed by 40 cycles of 95^{o}C for 10s and 60^{o}C for 30 sec. A melt curve program of 65-99^{o}C was included to check the specificity of qRT-PCR products. Three technical replicates per reaction were performed, and the mean Ct value was used for the analysis. A reverse transcription negative control (without reverse transcriptase) and a non-template negative control were included to confirm the absence of genomic DNA and check non-specific amplification. Using Primer Quest (www.idtdna.com/primer quest/ home/index) 20 primers were designed (Sheeja et al., 2015) as mentioned in Table 1. Candidate reference genes *viz*., EF1α (FP-GCTGACTGTGCTGTTCTCATTAT and RP-CTCGTGTCTGTCCATCCTTTGAA) described in chapter 1. Standard curves were generated for each gene with serial dilutions of pooled cDNAs (10-1 to 10-6) for calculating qRT-PCR efficiency. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) and transformed to a log 2 scale.

Table 1: Gene ontology analysis and qRT-PCR primers of TFs showing differential expression in Transcriptome

S. No	Contig ID	Fold change	Sequence description	Blast per. Identity	Blast top hit description	GO names	Forward primer (5' - 3')	Reverse primer (5' - 3')
1	647693	10.24	bHLH transcription factor [*Lilium* hybrid division I]	69.42%	*Musa acuminata* subsp. malaccensis transcription factor EGL1-like mRNA	XM_009385819.2	CTTGCTAATGGAAGGAACCAATG	CGGAGAATCTACTGAAGCATGT
2	587097	8.55	Putative WRKY transcription factor [*Musa balbisiana*]	81.98%	*Musa acuminata* subsp. malaccensis WRKY transcription factor WRKY51, transcript variant X2, mRNA	XM_009420281.2	TCGTATTCCAGCTCTCCCA	TCCGATTCAGGACGGAGAT
3	603014	6.29	bHLH transcription factor-like protein [*Musa paradisiaca*]	80.13%	*Musa acuminata* subsp. malaccensis transcription factor EGL1-like, mRNA	XM_009385819.2	CTACGATGGTTATCTGTGGAGAAA	CCTTTGTCCGATCTGTAGGTG
4	644375	6.19	transcription factor bHLH149-like [*Solanum lycopersicum*]	78.95%	*Musa acuminata* subsp. malaccensis transcription factor bHLH149-like, mRNA	XM_009385663.2	ATCCTCCTCTCCGGCAA	ACCCTCGTCTTCTTCTCCA
5	659408	5.82	transcription	85.44%	*Musa acuminata*	XM_009393878.2	CTTGTCACTTTCTG	CCTACTCTCCAA

			factor WRKY71 [Musa ABB Group]		subsp. malaccensis WRKY transcription factor WRKY71-like, mRNA		GCCATATT	GTTCTACGTC
6	668415	5.26	transcription factor bHLH30-like [*Fragaria vesca* subsp. vesca]	84.62%	*Ananas comosus* transcription factor bHLH30-like, mRNA	XM_020248645.1	CGCATCAACGGCCATCT	CCGCTTCAGCTCCTTCAC
7	647385	5.01	NAC transcription factor ONAC010-like [*Setaria italica*]	81.60%	*Musa acuminata* subsp. malaccensis NAC domain-containing protein 35 mRNA	XM_009383787.1	GATTGGGTCATGAACGAGTACC	CAGCGATGTAGCTTTCCTGTAT
8	677063	5.23	F-box and wd40 domain protein, putative	78.36%	*Musa acuminata* subsp. malaccensis uncharacterized WD repeat-containing protein alr3466, transcript variant X1, mRNA	XM_009418754.2	GCTTAAGCCTCGACGAGAAG	ATGGACTCGAGGCACTTAGA
9	510280	5.12	NAC domain protein [*Cymbidium* hybrid cultivar]	86.14%	*Musa acuminata* subsp. malaccensis NAC transcription factor 29-like,	XM_009412202.2	CGCCGACGTCGACATATAC	GGTGAAGAAGTACCACTCTCG

					mRNA			
10	647385	5.01	NAC transcription factor ONAC010-like [*Setaria italica*]	67.24%	*Musa acuminata* subsp. malaccensis NAC domain-containing protein 35, mRNA	XM_009383787.1	GATTGGGTCATGAACGAGTACC	CAGCGATGTAGCTTTCCTGTAT
11	663323	4.45	NAC domain-containing protein 29-like	77.37%	*Musa acuminata* subsp. malaccensis NAC domain-containing protein 83-like, mRNA	XM_009412290.2	CACAGCTTGAGGAAGAGAAAGA	AGAAGAGGACAGGTGGAGTAG
12	609334	10.08	WRKY31 (Fragment)	89.01%	*Musa acuminata* subsp. malaccensis WRKY transcription factor WRKY51, transcript variant X2, mRNA	XM_009420281.2	GAGCTACTGGTAAAGGAACTGG	CAGATTAAGACCTTCGTGGAGAA
13	489876	3.80	Putative WRKY transcription factor 39 [*Aegilopstauschii*]	84.77%	*Musa acuminata* subsp. malaccensis probable WRKY transcription factor 21, transcript variant X4, mRNA	XM_009415648.2	CCGGAGGGATATCAGATGATTTAT	AAATGGCAAATGTGCTACTACTG
14	658926	5.06	WRKY	79.80%	*Musa acuminata*	XM_009405246.2	TGCGATAGGTGCA	AGGAAGACGCTG

			transcription factor 54 [*Jatropha curcas*]		subsp. malaccensis probable WRKY transcription factor 41, mRNA		TCTGAAATA	AGGAAATG
15	609452	15.59	bHLH transcription factor [Lilium hybrid division I]	74.02%	*Nelumbo nucifera* transcription factor GLABRA 3-like, transcript variant X2, mRNA	XM_010243597.2	AGATATCACCATTGGCAGGAAG	CGATAGGGCCTTCTTTGGATAG
16	632760	5.15	transcription factor bHLH150-like [*Setaria italica*]	76.46%	*Musa acuminata* subsp. malaccensis transcription factor bHLH149-like, mRNA	XM_009411209.2	GCGTTCGCTCACCTTCTT	GCTGCAAGTTCCTCCTCAAA
17	660094	6.49	WRKY transcription factor 18 [*Oryza sativa* (japonica cultivar-group)]	73.02%	*Panicum hallii* WRKY transcription factor 55-like, mRNA	XM_025950044.1	CTACGATGGTTATCTGTGGAGAAA	CCTTTGTCCGATCTGTAGGTG
18	651973	4.84	putative bZIP transcription factor superfamily protein [*Zea mays*]	79.59%	*Durio zibethinus* bZIP transcription factor 18-like, mRNA	XM_022891899.1	ACATGATGTAACCTCCCATTAGAC	TCAGATGCATGAGGCACTAAAT

19	575876	4.47	Transcription factor HY5 (Protein LONG HYPOCOTYL 5) (bZIP transcription factor 56) (AtbZIP56)	86.44%	*Musa acuminata* AAA Group long hypocotyl 5 (HY5) gene, complete cds	MN394629.1	AGACTGAGCGGTGGACT	TGGAAGGCAAAGAGGGAATG
20	664145	4.78	Transcription factor bHLH61 isoform 1 [*Theobroma cacao*]	70.14%	*Camellia sinensis* transcription factor bHLH93-like, transcript variant X1, mRNA	XM_028267206.1	GTTCGAGGAGTTGGGTCTTC	CCTTGGGCTTCTCTTCCAATA

2.3.1.4. Mining of bHLH and WD40 transcription factors from whole genome

Two bHLH (603014 and 668415) and one WD40 TFs of length 289, 735 and 1027 bp respectively obtained from transcriptome were blasted against whole genome data base of turmeric. Sequences having high percentage of identity were selected. The deduced amino acid sequences of TFs were obtained by translating the cDNA sequences using the EMBOSS program TRANSEQ (Rice et al., 2000). Comparative sequence analysis TFs were performed online using blastp (Altschul et al., 1990). The open reading frame (ORF) was predicted by ORF Finder (Rombel et al., 2002).

2.3.1.5. Phylogenetic analysis

Phylogenetic analysis of sequences was performed using Clustal W (Thompson et al., 1997). program was used to align nucleotide sequences with the Gonnet scoring matrices. Phylogenetic analysis was performed with the MEGA7. Other plant bHLH such as (ZmLc)NP_001105339.2, (PhAN1) AAG25928.1, (VvMYC1) ACC68685.1, (FabHLH3) QIZ03068.1, (GL3) Q9FN69.1, (VvMYCA1) ABM92332.3, (MdbHLH33) ABB84474.1, (PhJAF13) AAC39455.1, (ZMIN1) AAB03841.1, and (AtTT8) NP_192720.2 and plant WD40 transcription factors (VvWDR1) ABF66625.2, (ZmPAC1) AAM76742.1, (PhAN11) AAC18914.1, (MdTTG1) ADI58760.1 TTG1, (FaTTG1) AFL02466.1, (AtTTG1) NP_001318637.1 were selected for phylogenetic tree. A rooted phylogenetic tree was generated using the maximum parsimony method employing anaheuristic search strategy. Gaps were treated as missing data. To determine relative level of support for the tree topology, bootstrap values were generated from 1000 replicates.

2.3.1.6. Molecular modelling

Structure homology was computed by the SWISS-MODEL (Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Switzerland) homology server (Waterhouse et al., 2018 and Bienert et al., 2017). The GMQE (Global Model Quality Estimation) and QMEAN parameter (Benkert et al., 2008) were determined to ascertain enough model quality. The model was validated using Ramachandran plot. Ramachandran plot was identified by Procheck program of Structural Analysis and Verification Server (Laskowski et al., 1993).

2.3.1.7. Estimation of curcumin content

Curcumin content of rhizome samples was analysed spectrophotometrically at 430 nm following the American Spice Trade Association (1968) procedure (Table 2).

Table 2: List of turmeric accessions used in the study with curcumin content

***Curcuma* Species**	**Source**	**Curcumin content (%)**
IISR Prathibha	ICAR-IISR germplasm	>5.5 (high)
Accession No. 200	ICAR-IISR germplasm	3.54 (low)
C. aromatica	ICAR-IISR germplasm	0.12 (very low)
Accession 449	ICAR-IISR germplasm	2.11% (low)

2.3.2. Functional characterisation of miRNAs

2.3.2.1 Plant material

To evaluate the expression of miRNAs regulating phenylpropanoid biosynthesis three candidates *viz.*, miR319, miR828 and miR858 were identified from RNA-Seq data from 4-month-old rhizomes of IISR Prathibha grown under different light regimes ((Red, Green, and White and open conditions served as control) reported to accumulate different levels of curcumin (Deepa et al., 2017, Deepa, 2018). The samples were collected from field experiments laid at the ICAR-Indian Institute of Spices Research (11.2994°N,75.8407°N), Kozhikode, Kerala (Chapter 1)

2.3.2.2. Mining of miRNA from miRNA transcriptome

miRNAs were retrieved from the miRNA deep transcriptome generated from Sandor Proteomics, Hyderabad, India using a high curcumin variety (IISR Prathibha) and a low curcumin accession (Accession 449) (Santhi et al., 2016) and miRNA specific primers were designed as per (Varkonyi-Gasic et al., 2007) Table 3.

Table 3: List of miRNA primers used for the study

Primer name	RT Primer (5'-3')	Forward primer (5'-3')
miR319d	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGAGCT	GCGGCGGTTTGGAATGAAGGG
miR828	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC TGGAAT	CGGCGGTCTTGCTTAAATGAGT
miR 858a	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC AAGGTC	GCGGCGGTTTCGTTGTCTGTTC
Universal RP	GTGCAGGGTCCGAGGT	
Primers used for the identification of stable reference gene		
18S rRNA	CCTTCCTCTAAATGATAAGGTTCAATG	GATTGAATGGTCCGGTGAAGTGTT
EF1α	GCTGACTGTGCTGTTCTCATTAT	CTCGTGTCTGTCCATCCTTTGAA
Ubiquitin	GCACTCTCGCTGACTACAAC	GGCTTGGTGTAGGTCTTCTTC
GAPDH	AACTGTAGCCCCACTCATTG	GCATCTTAGGGTATGTGGAGG
TUBULIN	GGCAGAGATCAGATGGTTCAG	TGGACAATGAAGCACTCTACG

2.3.2.3. End point PCR and expression profiling by qRT-PCR

For validation of three miRNAs *viz*., miR858, miR828 and miR319 identified from transcriptome data, total RNA was extracted from rhizomes of IISR Prathibha (Deepa et al., 2014) grown under different shade net conditions followed by removal of DNA using DNase I (Thermo scientific) at 37°C for 30 minute and inactivation by adding 1µl 50mM EDTA (Hi-media) at 65°C for 10 minutes. Primer design and RT-PCR was performed as previously reported (Varkonyi-Gasic et al., 2007) and resolved in 4% agarose (Sigma) gel under constant current of 90V till the bromophenol blue dye (Hi-media) front migrated to the bottom of the gel. The gel was visualized under UV using Gel Documentation System (Syngene). The gene expression was quantified by qRT-PCR for which firstly candidate reference gene was optimised and miRNAs screened under different light regimes (white, red, black, green and open conditions). Web-based tool RefFinder (RefFinder heartcure.com.au) and distribution of house-keeping genes is represented as box plot. qRT-PCR was done to evaluate the relative expression of the three miRNAs, using QuantiFast SYBR Green PCR kit (Qiagen) on Rotor Gene-Q real time PCR (Qiagen). All the reactions were performed three times. qRT-PCR parameters were 95°C for 5min, 40 cycles of 95°C for 10s, 60°C for 30s. Melt curve analysis was performed at 70-95°C to check the specificity of PCR products. Relative gene expression was calculated using $2^{-\Delta\Delta Ct}$ method. 18S rRNA was used as the endogenous control.

2.3.2.4. Estimation of curcumin content as mentioned in Chapter 1, Section 1.3.2.

2.4. RESULTS AND DISCUSSION

2.4.1. Identification of TFs from *Curcuma* transcriptome

Transcriptome analysis is widely reported as an effective method for identifying regulatory molecules like TFs in many studies involving phenylpropanoids (Zong et al., 2019; Song et al., 2019; Todd et al., 2010). Aligning the annotated transcripts of *C. longa* and *C. aromatica* to the AGRIS database resulted in the identification of unigenes belonging to 39 transcription factor families in *C. longa* and 36 in *C. aromatica*, which included bHLH, MYB, WRKY, MADS, NAC, bZIP, ARR-B, C2H2 and AP2-EREBP TFs (Sheeja et al., 2015). Among these, seven bHLH (contig ID nos. 647693, 603014, 644375, 644375, 668415, 609452, 632760), six WRKY (contig ID

nos. 587097, 659408, 658926, 660094, 609334, 489876), three NAC (contig ID nos. 647385, 510280, 663323), two bZIP (contig ID nos. 651973, 575876) and one WD40 (contig ID no. 677063) showed maximum fold change *vis-a-vis* curcumin (Table 2). Several families of TFs have been described to be regulators of plant secondary metabolism, which include bHLH (Hong et al., 2012; Shoji et al., 2011; Zhang et al., 2011; Todd et al., 2010), WD40 (Li et al., 2014), MYB (Teng et al., 2005, Gonzalez et al., 2008; Yamagishi et al., 2010; Xu et al., 2017), WRKY, (Suttipanta et al., 2011; Ma et al., 2009), bZIP, (Yu et al., 2012), NAC (Saga et al., 2012) (Table 1).

2.4.2. Expression pattern of bHLH and WD40 TFs vs curcumin

Biosynthesis of phenylpropanoid derivatives like flavonoids is generally controlled by MYB TFs and MYB-bHLH-WD40 (MBW) complexes (Schaart et al., 2013; Li et al., 2014; Xu et al., 2015). Analysing qRT-PCR expression patterns of TFs is a common strategy to validate the RNA-Seq data (Foong et al.,2020; Xia et al., 2020). A similar study was conducted in *Hibiscus cannabinus* to evaluate the transcription factor families associated with kenaf response drought stress. They identified different TFs such as AP2/ERF, MYB, NAC, and WRKY families (An et al., 2020). Twenty shortlisted TFs from transcriptome analysis were further evaluated by qRT-PCR with IISR Prathibha as control. Among those, contig IDs 603014 (Fold change-2.05 in Acc.200 and 2.58 in *C. aromatica*) and 668415 (bHLH TFs) (Fold change-32.44 in Acc.200 and 115.36 in *C. aromatica*) and contig ID 677063 (WD40 TF) (Fold change-3.22 in Acc.200 and 66.71 in *C. aromatica*) showed the highest relative fold change with respect to curcumin (low vs high) (Figure 2).

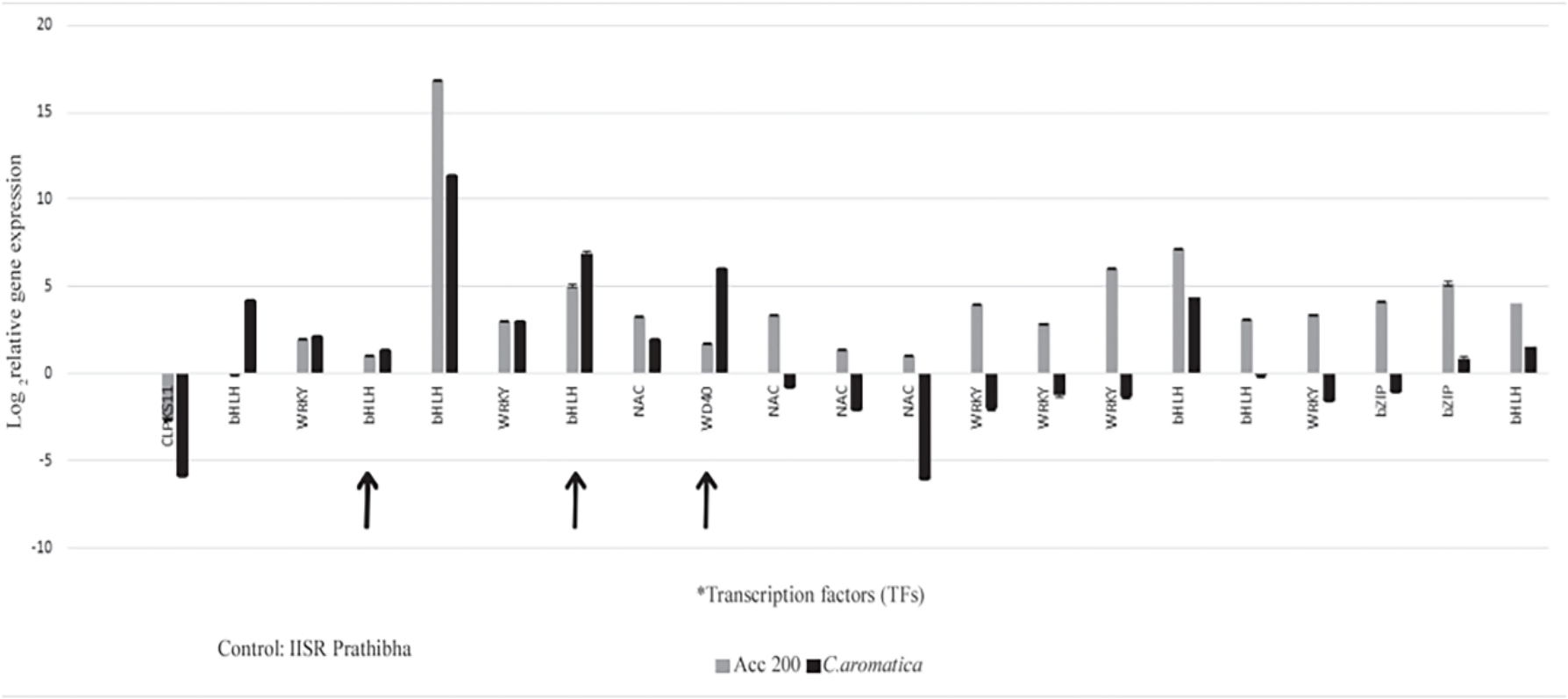


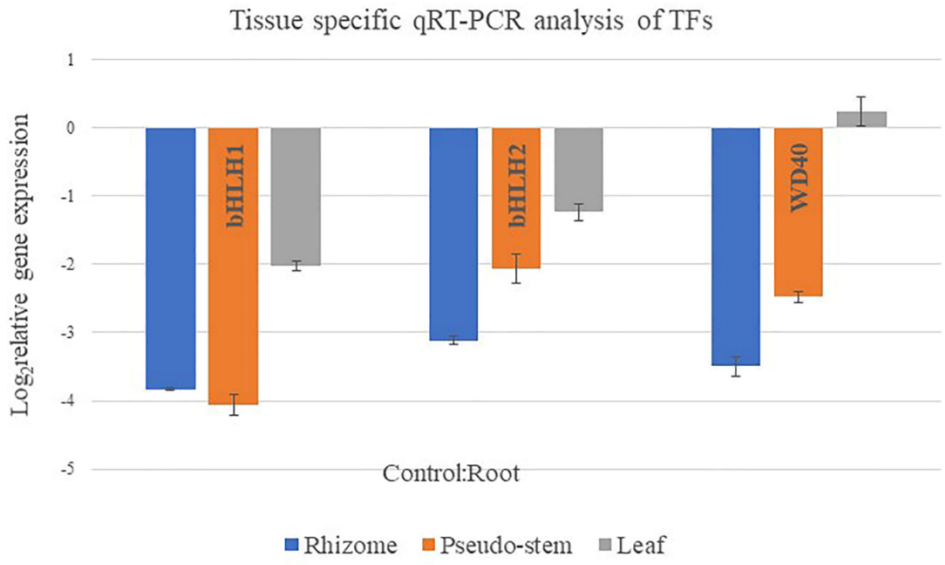
Figure 2: Mean normalized expression of TFs mined from transcriptome. Error bars are standard errors of the mean from three technicalreplicates. *TFs are in the same order as in Table 1, the arrow represents the novel candidate TFs identified by qRT-PCR.

The expression of *ClPKS11*, upregulated under high curcumin conditions (Deepa et al., 2017), was high in IISR Prathibha while it was downregulated in *C. aromatica* and Accession No. 200. A negative correlation was observed in case of bHLH TFs (Contig IDs 603014 and 668415) and WD40 TF (contig ID 677063) with respect to *ClPKS11* expression (Fig. 4). It is a usual practice that if more than one candidate gene is identified to correlate with biosynthesis of the key metabolite, these genes are employed as baits to select out other novel candidates of the pathway based on FPKM and expression analysis data. Several studies have reported that bHLH members play a central role in transcriptional regulation of biosynthetic pathways by forming a complex with WD40 and MYB TF proteins (Zhao et al., 2019). TF regulators that interact with bHLHs or R2R3-MYBs organize or disrupt MBW complex formulation, regulating flavonoid production (Li et al., 2014). The qRT-PCR expression trends were in congruence with fold change values in RNA-seq data. Three novel TFs belonging to bHLH and WD40 were found to show binding sites with several MYB TFs identified in our earlier studies (Sheeja et al., 2015). Blastn search indicated the unique status of all three TFs based on the per cent identity (Table 1). 603014 (bHLH) showed that 80.13% identity with *Musa acuminata subsp. malaccensis* transcription factor EGL1-like, mRNA. 668415showed 84.62% similarity with transcription factor bHLH30-like [*Fragaria vesca* subsp. vesca] and 677063 showed that 78.36% identity with *Musa acuminata* subsp. malaccensis uncharacterized WD repeat-containing protein alr3466, transcript variant X1, mRNA. novel TFs, along with already identified MYB TFs will be instrumental in future studies to confirm the mechanism of the molecular regulation of curcumin biosynthesis. Being novel, two bHLH genes (Contig ID 603014 and 668415) were designated as *ClbHLH1*, *ClbHLH2* and the WD40 Contig (ID 677063) was termed as *ClWD40*.

The expression patterns of the bHLH and WD40 TFs evaluated in spatio-temporal tissue specific and nutrient management studies further indicated the involvement of these TFs in pathway regulation. Tissues like rhizome, root, pseudo-stem and leaves when evaluated with root as control indicated that *ClbHLH1* expression was high in leaf (fold change- 0.243) followed by rhizome (fold change- 0.07) and pseudo-stem (fold change- 0.062). In case of *ClbHLH2* and *ClWD40* high expression was seen in leaf (fold change- 0.353 and 2.29 respectively) followed by pseudo-stem (fold change- 0.244 and 0.143respectively) and rhizome (fold change-

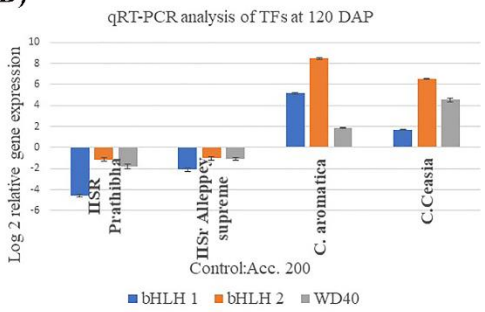
0.122and 0.082respectively). Compared to rhizomes *ClbHLH1*, *ClbHLH2* and *ClWD40* showed higher expression in leaves (Figure 3 (A)).

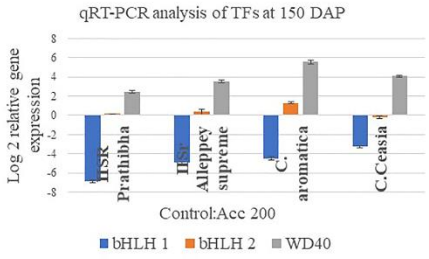
(A)

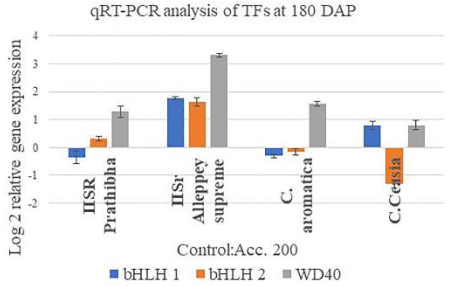


Different locations	Curcumin content(%)
IISR Prathibha rhizome	5.37

(B)

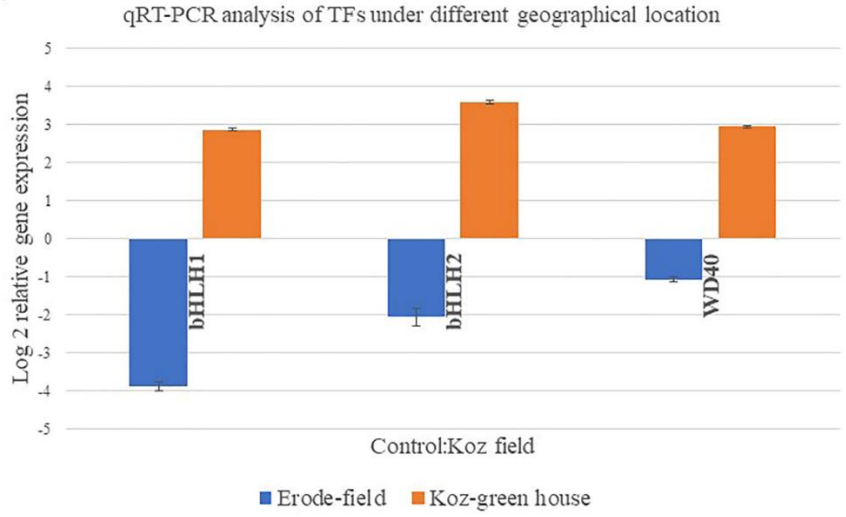






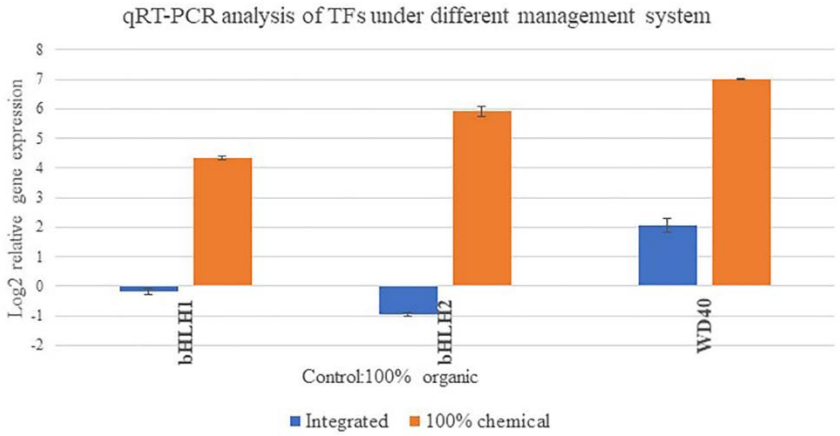
Different accessions	Curcumin content(%)		
	120 DAP	150 DAP	180 DAP
IISR Prathibha	5.37	4.99	3.09
Alleppey supreme	5.21	5.32	4.76
Acc.200	3.54	1.74	3.09
C. aromatica	0.12	0.12	0.48
C.ceasia	0.16	0.04	0.07

(C)



Different locations	Curcumin content(%)
Koz-field	5.37
Erode-field	3.68
Koz-green house	3.01

(D)



Type of management	Curcumin content(%)
100% organic	5.71
Integrated	6.15
100% chemical	4.79

Figure 3: Relative expression pattern of *ClbHLH1*, *ClbHLH2* and *ClWD40*. (A) different tissues, (B) different developmental stages, (C) different geographical location, (D) different management studies

The temporal expression analysis of three TFs when done using Acc. 200 as control indicated that expression of *ClbHLH1* (fold change- 0.05, 0.012, and 0.79 in 120 DAP, 150 DAP and 180 DAP IISR Prathibha; 17.50, 2.29, 1.07 respectively in *C. aromatica*), *ClbHLH2* (fold change- 0.75, 1.14, and 2.29 in 120 DAP, 150 DAP and 180 DAP IISR Prathibha; 130.68, 2.29, and 1.07 respectively *C. aromatica*) and *ClWD40* (fold change- 0.920, 4.08, and 2.11 in 120 DAP, 150 DAP and 180 DAP IISR Prathibha; 4, 29.85, and 2.828 in 120 DAP, 150 DAP and 180 DAP *C. aromatica*) negatively correlated with curcumin in different accessions at 120 DAP

However, expression did not correlate with curcumin in case of *C. aromatica* and *C. ceasia* that do not synthesise curcumin, where a higher gene expression was observed. These results indicated that these TFs might be dedicated candidates involved in regulation of curcumin in systems actively synthesising curcumin and probably these genes could be involved in alternate pathways in those species. Expression analysis of samples collected from 150 DAP showed that *ClWD40* (fold change-4.08) and *ClbHLH2* (fold change-1.14) increased slightly in high curcumin accessions and expression of *ClbHLH1*(fold change-2.29) was downregulated in low curcumin accessions. qRT-PCR analysis of rhizomes at 180 DAP showed that expression of all the transcription factors was high (fold change-0.79, 2.29, and 2.11 in high curcumin and 1.07,1.07 and 2.82) in high curcumin accessions and low in low curcumin accessions. Curcumin content was reduced when rhizomes reached 150 DAP (4.99%) and further reduced at 180 DAP (3.09%). The expression of three transcription factors correlated with curcumin content in these experiments too. But the expression of all three TFs reduced in low curcumin accessions at 150 DAP and further reduced at 180 DAP (Figure 3(B)).

Evaluation of impact of different geographical locations on TF expression was done involving samples collected from Koz- field having high curcumin content (>5.5%), followed by Erode field (3.68%) and Koz-greenhouse (3.01%). It was observed that expression of *ClbHLH1*, *ClbHLH2* and *ClWD40* was high (0.06, 0.203, and 0.5 in Koz-field; 7.46,12.12, and 6.72) in Koz-green house where curcumin content was very low with Koz-field kept as control (Figure 3(C)).

Experiments on effect of TFs on curcumin biosynthesis under 100% organic, integrated and 100% chemical management practices using IISR Prathibha indicated that expression of TF varied with environment *vis-a-vis* curcumin. In integrated management system the expression of *ClbHLH1*, *ClbHLH2* and *ClWD40* was low (fold change-0.80,

0.53 and 3.73 respectively) where curcumin content is high in 100% chemical expression of three TFs high (fold change-17.14, 48.5 and 84.44 respectively) where curcumin content is low. The expression of TFs was higher under chemical management conditions compared to organic and integrated systems. Integrated management practices yield turmeric with high curcumin content of 6.15% compared with those grown under organic (5.71%) and 100% chemical management systems (4.79%) (Figure 3(D)). The study also highlights the importance of integrated/organic management system over chemical management system. Thus, the study also indicates that integrated/organic farming practices are better than chemical management for maximizing gene expression and curcumin content. The expression of all the TFs were high in 100% chemical practices where curcumin content was low (Figure 3(D)). Hence from this qRT-PCR analysis it is observed that all three TFs were negatively correlated with curcumin under the experimental conditions, and this also indicates that these transcription factors are putative negative regulators of curcumin biosynthesis.

2.4.3. Full length gene mining of two bHLH and WD40 from whole genome

Through local blast, nucleotides of 1585 bp, 2039 bp and 1425 bp were retrieved from the whole genome data (Chakraborty et al., 2021) that showed more than 90% identity with transcriptome derived sequences of *ClbHLH1*, *ClbHLH2* and *ClbHLH3*. ORFs of those transcription factors were found to be of 1071bp, 753bp and 1425bp as obtained from ORF finder. The predicted amino acid sequence of *ClbHLH1* was highly similar to the reported plant *Zingiber officinale* transcription factor EGL1-like protein (XM_042539192.1). The predicted amino acid sequence of *ClbHLH2* was highly similar to the reported plant *Zingiber officinale* transcription factor bHLH30-like isoform X3 (XP_042450739.1), while predicted amino acid sequence of *ClWD40* showed high similarity to the plant WD repeat-containing protein 86-like [*Musa acuminata subsp.*malaccensis] (XP_009417029.1).

2.4.4. Phylogenetic tree analysis

Phylogenetic analysis showed that *ClbHLH1* clustered with *Arabidopsis thaliana* bHLH (AtTT8) transcription factor. But *ClbHLH2* did not cluster with any of the existing plant bHLH factors belonging to the MBW complex indicating its novel status. *ClWD40* clustered with the *Arabidopsis thaliana* WD40 (AtTTG1) transcription factor showing a similarity. So it could be presumed that ClbHLH1 and ClWD40 transcription factors together with MYB form a MBW ternary complex that regulate the structural

genes involved in curcumin biosynthesis pathway. It could also be assumed that ClbHLH2 could be an independent regulator of the structural genes (Figure 4 (A) and (B)).

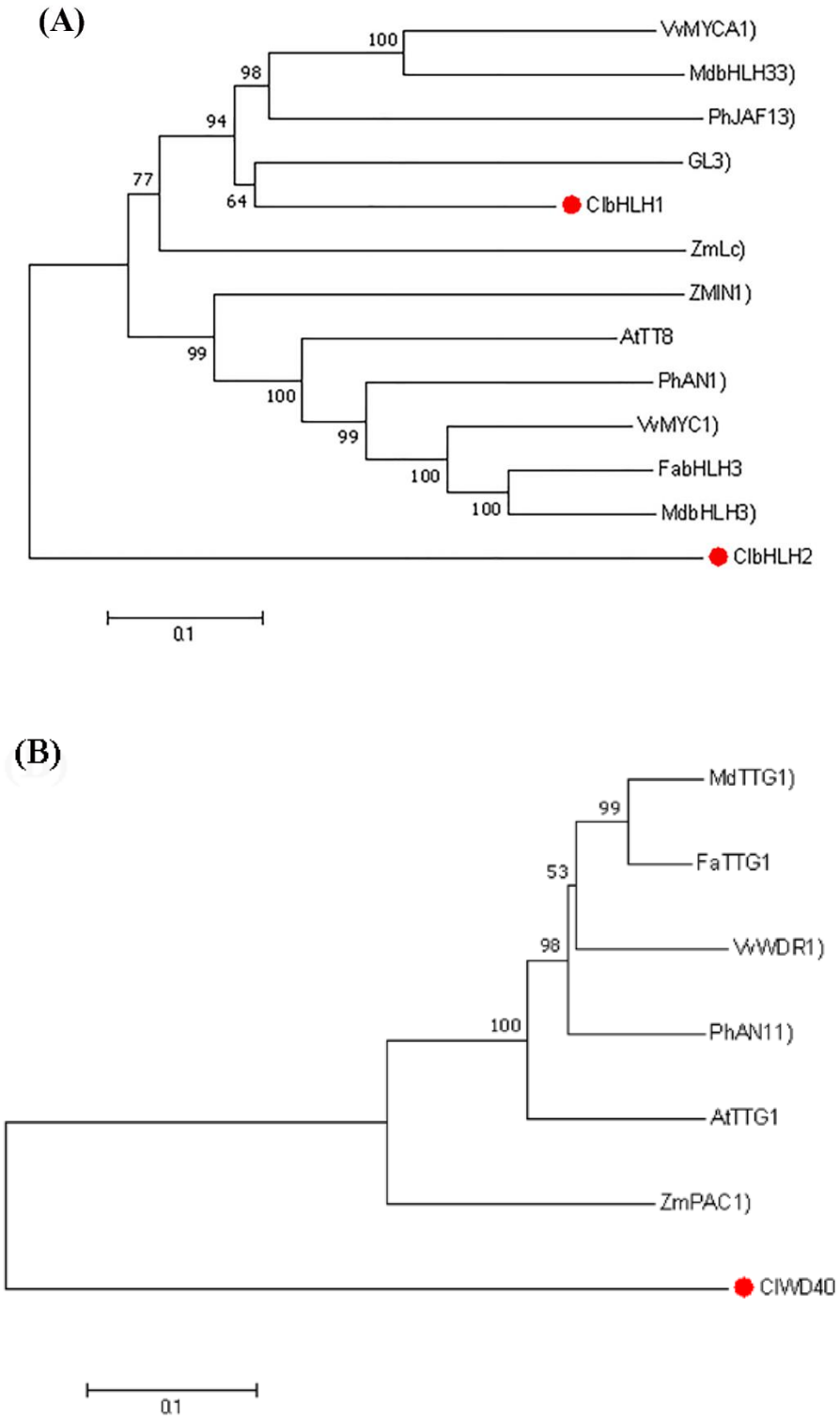


Figure 4: Phylogenetic tree analysis of (A) ClbHLH1, ClbHLH2 and (B) ClWD40

2.4.5. Molecular modelling

By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction of ClbHLH1 revealed that α-helices (37.92%) were the main structural elements, and random coils (35.11%) and extended strand (18.82%) were scattered in the entire protein (Figure 5 (A). By using the known *Arabidopsis thaliana* Transcription factor MYC3 showing (30.32%) identity (4ywc.2.A) (Ritter and Schulz, 2004) as a template in the SWISS-MODEL program (Arnold et al., 2006), The 3D structure of ClbHLH1 protein model (Figure 5 (B)) was confirmed using Ramachandran Plot statistics (Figure 5(C)).

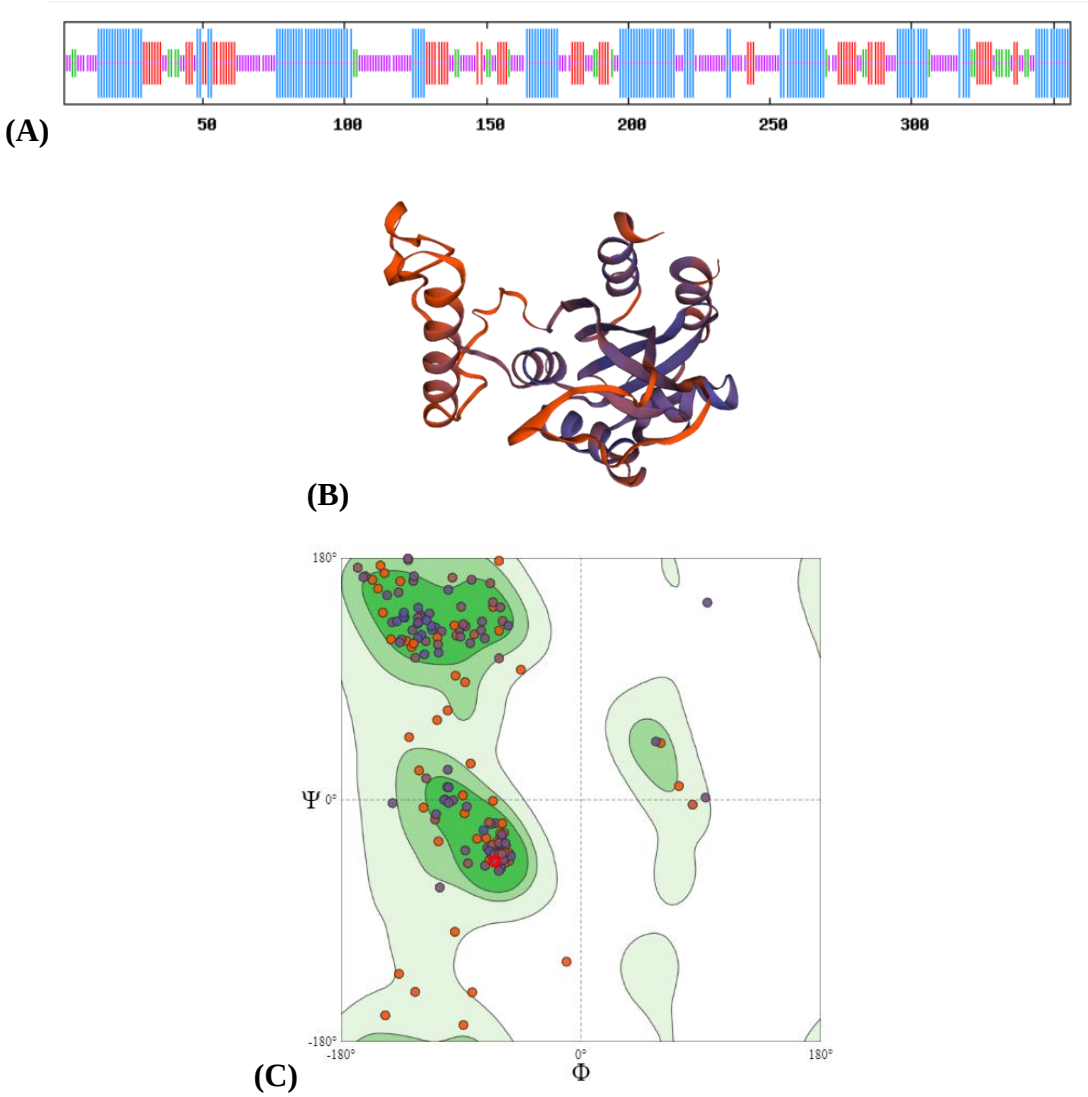


Figure 5: Secondary structure (A)Molecular modelling of *C. longa* ClbHLH1 (B) 3-D structure and (C) Ramachandran plot determined by PROCHECK

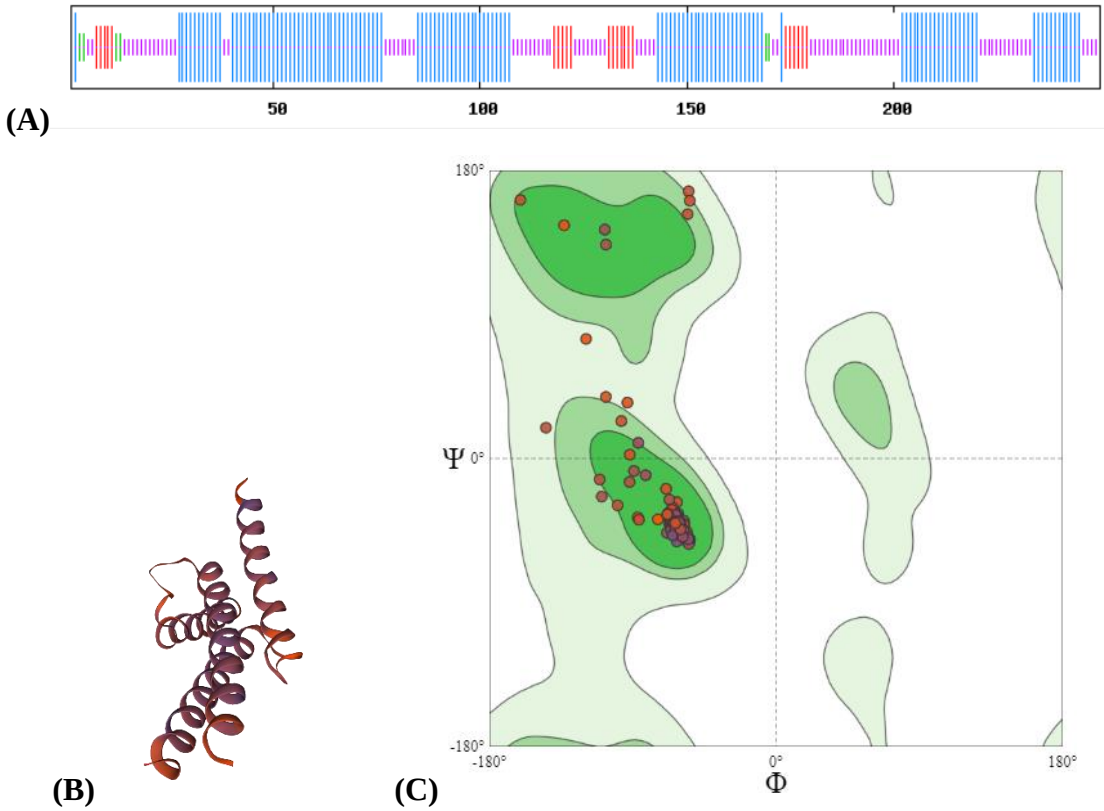


Figure 6: Secondary structure (A)Molecular modelling of *C. longa* ClbHLH2 (B) 3-D structure and (C) Ramachandran plot determined by PROCHECK

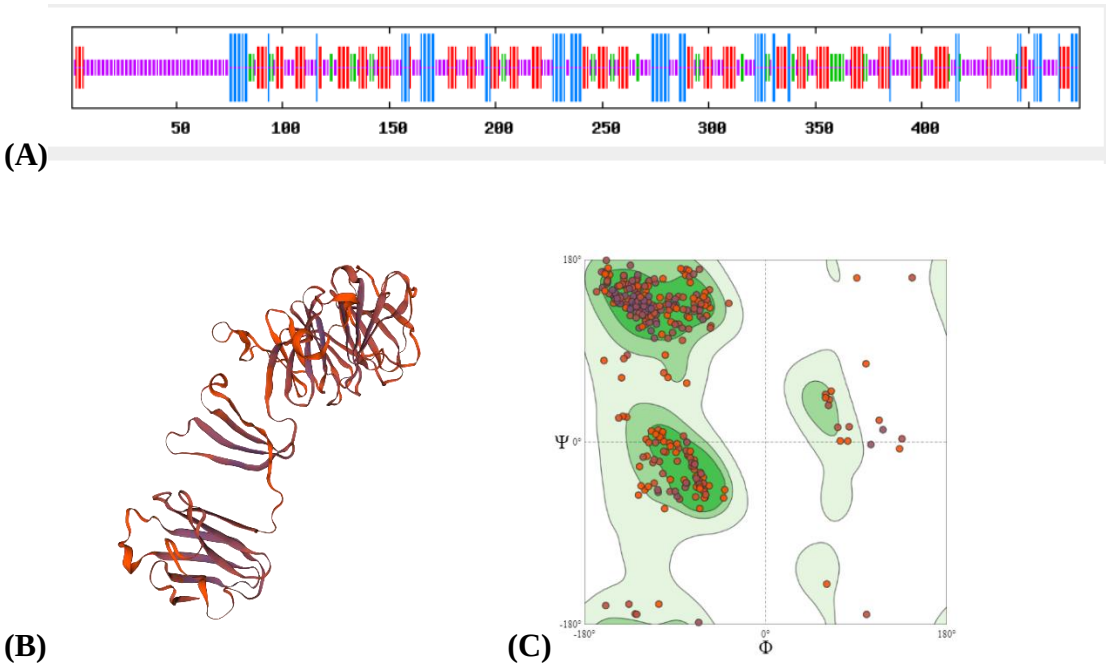


Figure 7: Secondary structure (a)Molecular modelling of *C. longa* ClWD40 (b) 3-D structure and (c) Ramachandran plot determined by PROCHECK

By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction of *ClbHLH2* revealed that α-helices (52.80%) were the main structural elements, and random coils (35.60%) and extended strands (9.20%) were scattered in the entire protein (Figure 6 (A). By using the known *Arabidopsis thaliana* Transcription factor MYC2 (5gnj.3.A) (Ritter and Schulz, 2004) as a template in the SWISS-MODEL program (Arnold et al., 2006). The 3D structure of ClbHLH2 protein model (Figure 6 (B)) was confirmed using Ramachandran Plot statistics (Figure 6 (C)).

The predicted amino acid sequence of ClWD40 was highly similar to the reported plant WD repeat-containing protein 86-like [*Musa acuminata* subsp. malaccensis] (XP_009417029.1). By using the SOPMA program (Geourjon and Deléage, 1995), the secondary structure prediction of ClWD40 revealed that random coil (47.89%), extended strand (27.85%) and α-helices (15.40%) were the main structural elements that were scattered in the entire protein (Figure 7 (A). By using the known WD repeat-containing protein 36 Cryo-EM structure of the human SSU processome, state pre-A1 MYC2 (7mq8.25.A) (Ritter and Schulz, 2004) as a template in the SWISS-MODEL program (Arnold et al., 2006). The 3D structure of ClWD40 protein model (Figure 7 (B)) was confirmed using Ramachandran Plot statistics (Figure 7 (C)).

2.4.6. Identification of conserved miRNAs

Twenty-nine conserved miRNAs could be identified from small RNA libraries of the low curcumin (2.11%) Accession 449 and a high curcumin (>5%) variety IISR Prathibha and the relative gene expression was correlated with curcumin content. On the basis of deep sequencing data, miRNAs with fold expression greater than 2.0 were chosen that includes 14 miRNAs which were down-regulated and 15 miRNAs that were upregulated. The expression levels of miR168b, miR167a, miR159q, miR156e miR164d, miR396e and miR396h were significant. These miRNAs could be involved in regulation of biosynthesis and need to be confirmed through expression studies.

2.4.7. Expression profiling of miRNA under different light regimes

We have already identified and shortlisted about 20 MYBs from transcriptome studies conducted involving samples under contrasting curcumin contents (Unpublished). MYBs are TFs that are normally downregulated by small RNA molecules like miRNAs. These results encouraged us to check the miRNA mediated regulation of MYB TFs *vis-a-vis* curcumin content. IISR Prathibha grown under different shade nets showed a

differential accumulation of curcumin based on the shade net colour (Deepa et al., 2017). Curcumin content under green (4.93%), white (4.84%), red (4.45%) and open conditions (5.22%) were evaluated and correlated with that of miRNAs. Three conserved miRNAs *viz.*, *miR319* (Santhi et al., 2016; Santhi 2018), *miR828* (Deng and Lu, 2017), *miR858* (Sharma et al., 2016) reported to play a big role in phenylpropanoid metabolism through regulation of MYB transcription factors were amplified under the above experimental conditions (Table 3). All the three miRNAs could be amplified *via* end point PCR (Figure 8, 9 and 10) as well as real time PCR (Figure 11,12, and 13). However, only miR319 showed a negative correlation under all three experimental conditions with maximum expression under red shade net and lowest under green shade net conditions. It could be presumed that miRNA 319 could probably be an inhibitor of MYBs that are activators of curcumin biosynthesis. A similar work conducted by (Thiebaut et al., 2012), on miR319 also reports the targeted cleavage of the gene GAMYB during cold stress in sugar cane. However, more studies are required to confirm the results.

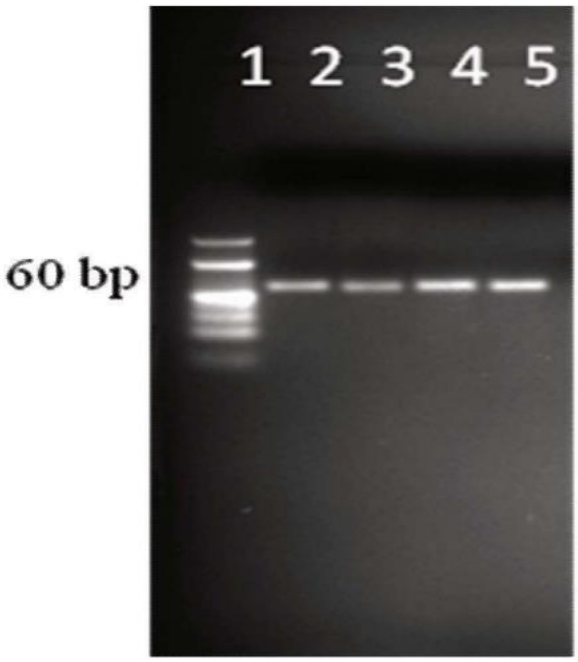


Figure 8: Amplification of miR319 in rhizomes under different light regime: Lane 1-10bp ladder, Lane 2-Control(open), Lane 3-Red shade net, Lane 4-Green shade net, Lane 5-White shade net.

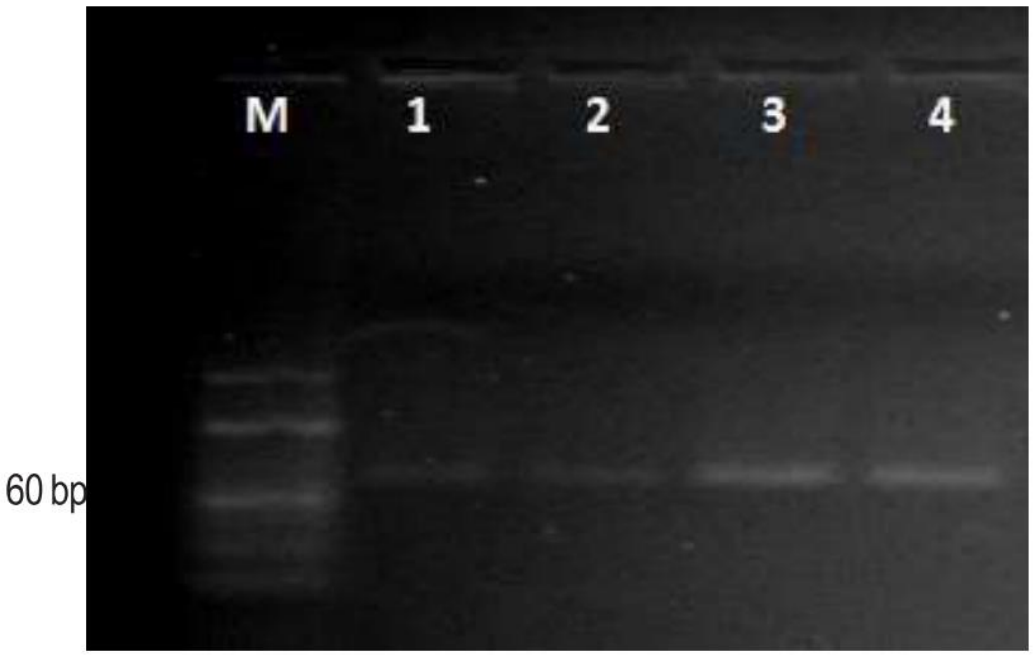


Figure 9: Amplification of miR828 in rhizomes under different light regime: Lane 1-10bp ladder, Lane 2-Control(open), Lane 3-Red shade net, Lane 4-Green shade net, Lane 5-White shade net.

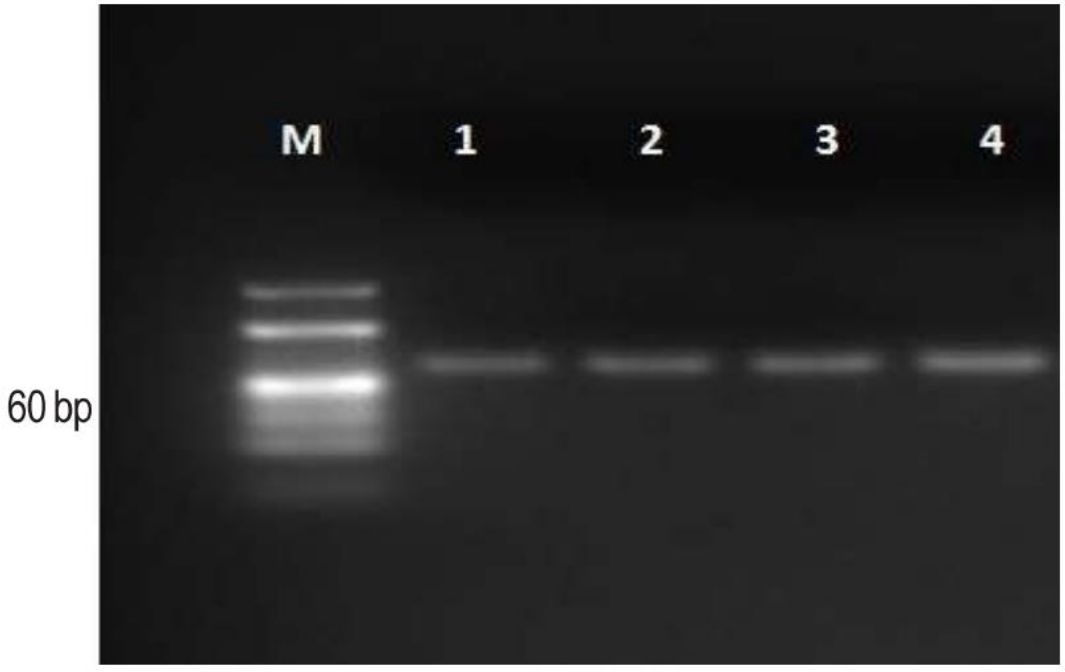


Figure 10: Amplification of mi858 in rhizomes under different light regime: Lane 1-10bp ladder, Lane 2-Control(open), Lane 3-Red shade net, Lane 4-Green shade net, Lane 5-White shade net.

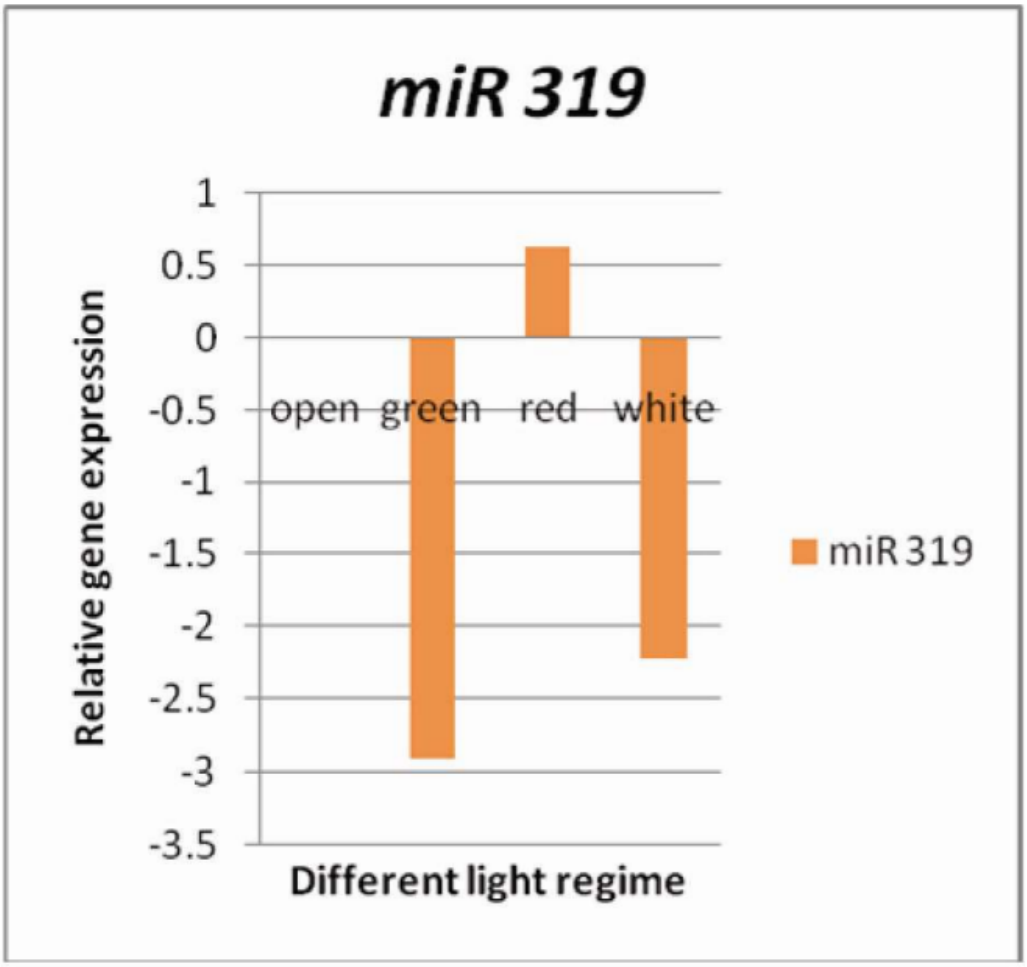


Figure 11: qRT-PCR analysis of miR319 under different light regime, control: IISR Prathibha grown in open condition

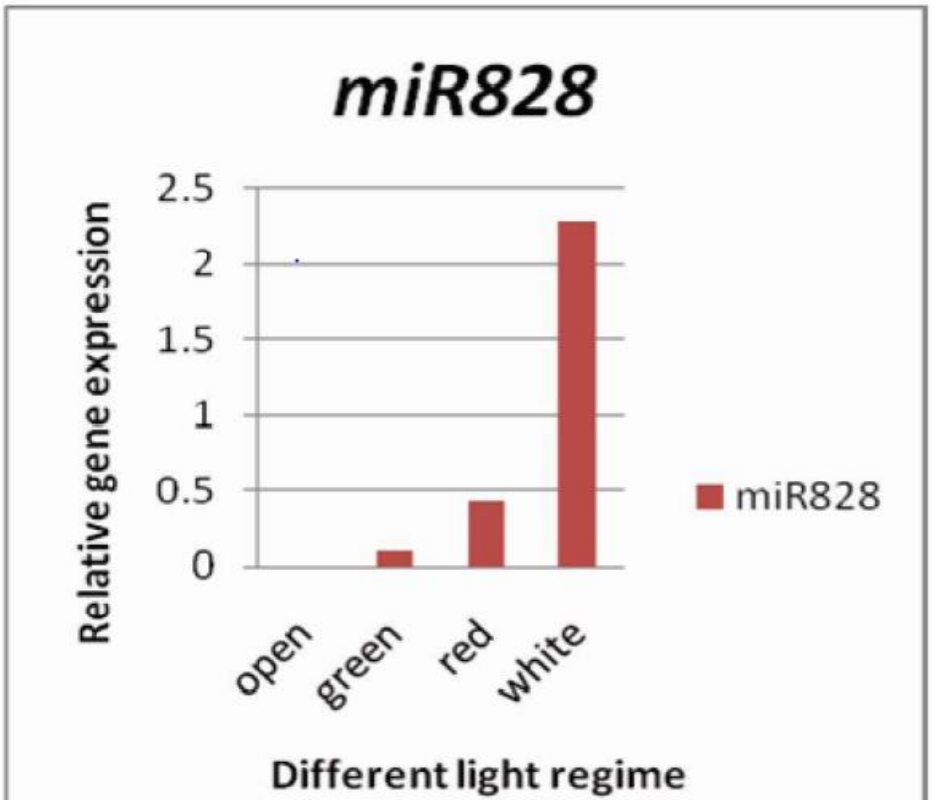


Figure 12: qRT-PCR analysis of miR828 under different light regime, control: IISR Prathibha grown in open condition

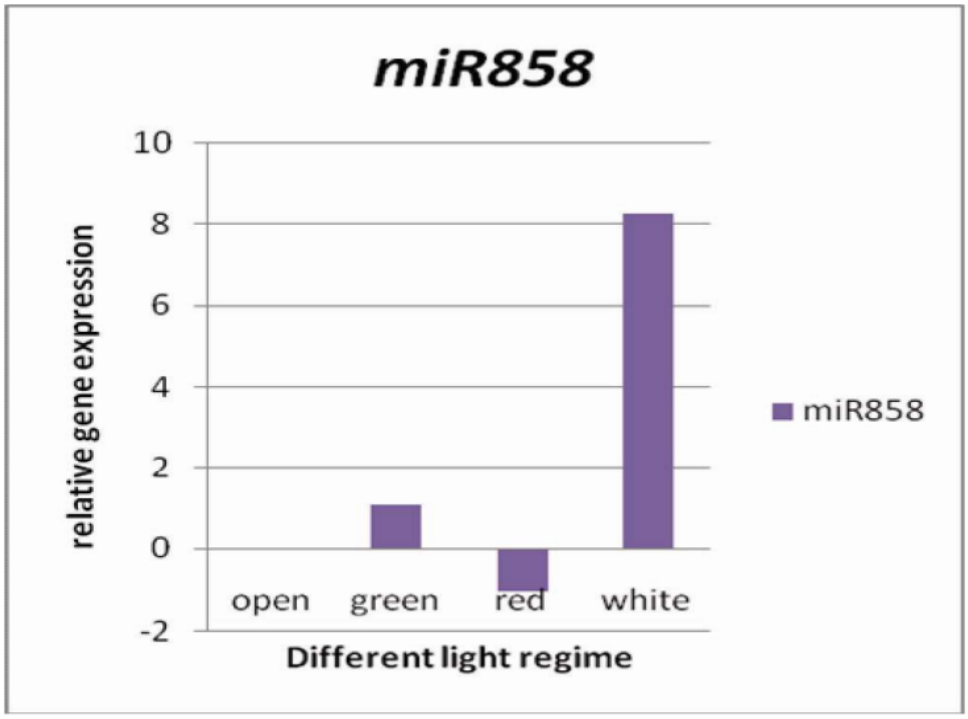


Figure 13: qRT-PCR analysis of miR858 under different light regime, control: IISR Prathibha grown in open condition

2.5. CONCLUSION

Curcuminoids are phenylpropanoid derivatives and the biosynthesis pathway is controlled by several transcription factors (TFs). bHLH, WD40 and MYB TFs could be the most important TFs regulating curcumin biosynthesis in turmeric as evident bt the presence of 20 TFs that showed differential expression in comparative transcriptomes under contrasting curcumin. Among these, two bHLH and one WD40 TFs showed maximum comparative fold change and negative correlation *vis-à-vis* curcumin in qRT-PCR based co expression analysis. The results of comparative transcriptome and qRT-PCR analysis were in congruence, indicating their putative role as negative regulators. The characteristics of the TFs by full length gene mining, phylogenetic tree analysis and molecular modelling confirmed their regulatory capacities and protein structures.

MicroRNA are a class of small endogenous non coding RNAs (sRNA) that could play a major role in gene regulation. miRNA mediated regulation of biosynthesis of curcumin is an unexplored area. We have identified through transcriptome analysis, about 29 miRNAs that showed differential expression with respect to curcumin in turmeric accessions with contrasting curcumin content. Expression of one of the conserved miRNAs viz., miR319 showed a negative corelation to curcumin when plants were grown under different light regimes favouring differential curcumin accumulation in rhizomes. This miRNA is a potential candidate for further studies on regulation of biosynthesis of curcumin.

Publication/abstract

- **Prashina Mol. P.,** Aparna, R. S., Sheeja, T. E., and Deepa, K., (2021) Novel bHLH and WD 40 transcription factors from turmeric (*Curcuma longa* L.) as putative regulators of curcumin biosynthesis. Journal of Plantation Crops, 49(1) 20-27.
- TE Sheeja, R Santhi, K Deepa, **P. Prashina Mol**, R. S. Aparna, A. Giridhari (2018). Uncovering roles of microRNA in regulation of curcumin biosynthesis in turmeric (*Curcuma longa* L.). *International journal of Innovative Horticulture*, 7(2), 146-149.

Chapter 3

iTRAQ BASED PROTEOMIC APPROACHES FOR VALIDATION OF PATHWAY GENES FOR CURCUMIN

3.1. INTRODUCTION AND OBJECTIVE

The proteomic approach provides a thorough estimation of proteins induced in response to a certain treatment or environment, and these studies provide both quantitative and active information of protein profiles, as well as protein-protein interactions (Anderson et al.,1998). Currently, functional genomic strategies such as gene microarray and RNA sequencing (RNA-Seq) are based on transcriptional (mRNA) level data. However, it is clear that transcriptional approaches alone cannot effectively reflect post-transcriptional/translational level alterations as the correlation between expressive abundance of genes and proteins is generally feeble under the same experimental conditions (Liu et al., 2016a), particularly when the expressive abundance of proteins is very low. Furthermore, gene expression patterns may not accurately depict the complex post-translational modifications, sub-cellular localization/transfer of proteins, and protein-protein interactions. Therefore, direct study of protein is adopted by many workers to gain a reliable understanding on the active patterns at protein level (Wang et al., 2016), which may be authenticated using complementary approaches. Isobaric tags for relative and absolute quantification (iTRAQ) are commonly used in discovery-based proteomics (Noirel et al., 2011; Treumann and Thiede, 2010). At present, iTRAQ stands out from traditional proteomic technologies because of its unique advantages, such as accurate quantification, high-efficiency sample separation, a high identification rate, and excellent instrument performance (Evans et al., 2012). iTRAQ analysis has been further strengthened by powerful bioinformatics tools and statistical analyses (Herbrich et al., 2013). Isobaric tags for relative and absolute quantification (iTRAQ) are employed widely with proven value in discovery-based proteomics (Noirel et al., 2011; Treumann and Thiede, 2010), which allows for simultaneous protein identification and quantification obtained by tandem mass spectrometry (MS/MS) from peptide fragments and low mass reporter ions, respectively (Evans et al., 2012). Since the tags are isobaric, the signal in the mass spectrometer is the sum of the peptide contribution from all samples, so there is a gain in sensitivity (Evans et al., 2012). As a part of the global proteomics analysis approach, we utilized isobaric tags for relative and

absolute quantification (iTRAQ) technique, followed by the identification and quantification of tagged peptides with liquid chromatography-tandem mass spectrometry (LCMS/MS). iTRAQ based proteomic analyses were employed to study the differences in protein expression profiles in high curcumin condition, low curcumin condition and very low curcumin condition in the IISR Prathibha (CLP), Acc. 200 (CLA) and *C. aromatica* (CA). This is the first report that focuses on investigation of the proteome related to curcumin biosynthesis in *Curcuma longa*. In this study genes identified from RNA-Seq data was validated using iTRAQ in samples of different curcumin status. We previously identified different structural genes and regulatory genes involved in curcumin biosynthesis using RNA-Seq. Here we have attempted to validate the candidate genes putatively involved in curcumin biosynthesis using iTRAQ.

3.2. REVIEW OF LITERATURE

iTRAQ (Isobaric Tags for Relative and Absolute Quantification) has attracted growing interest and is used in proteomics research as a technology that permits simultaneous quantification of proteins in various samples. Many different techniques have been developed to simultaneously compare protein levels across multiple samples. One method that has gained increased attention is iTRAQ (Karp et al., 2010), a shotgun technique that uses Isobaric Tags for Relative and Absolute Quantitation. Compared to other methods such as 2DE (O'Farrell, 2010), ICAT (isotope-coded affinity tags) (Gygi et al., 1999), and DIGE (differential gel electrophoresis) (Patton et al 2002), iTRAQ offers improved quantitative reproducibility, higher sensitivity (Wu et al., 2006) and has broad applications in proteomics research (Casado-Vela et al., 2010; Kilner et al., 2011; Skorobogatko et al., 2011). By using four or eight isobaric tags, iTRAQ can simultaneously analyze up to eight biological samples (Choe et al., 2007; Ross et al., 2004). The four reagents used in the 4-plex version of iTRAQ are named 114, 115, 116 and 117 along with four additional reagents 113, 118, 119 and 121. Each reagent is composed of a peptide reactive group and an isobaric tag that consists of a reporter group and a balance group. The peptide-reactive group specifically reacts with primary amine groups of peptides. The reporter group gives strong signature ions in tandem mass spectrometry (MS/MS) and is used to determine the relative abundance of a peptide. The balance group keeps the overall mass of the isobaric tag constant. With this property, identical peptides labelled with different isobaric tags will not be distinguishable in mass spectrometry. In the experimental workflow for iTRAQ, unlabeled protein samples are

first trypsin-digested and labelled with different isobaric tags independently. These labelled peptides from different samples are then mixed together and separated by liquid chromatography. Identical peptides from different samples labelled with different isotopes are chromatographically indistinguishable and appear as a single precursor. The isolated peptides are finally run through MS/MS for further fragmentation to generate a collection of mass spectra. The property of isobaric tags allows otherwise identical peptides from different samples to be detected as a single peak by mass spectrometry and to produce a single set of sequencing ions in MS/MS. The ion signals produced from the reporter regions together with the normal fragment ions provide information on peptide identification and quantitation for different samples. Using softwares such as MASCOT (Matrix Science Inc., Boston, MA, USA), a protein database search can be performed on the fragmentation data to identify the labelled peptides and hence the corresponding proteins. The relative abundance of low molecular mass reporter ions generated from the isobaric tags can then be used to quantify the relative abundance of peptides and proteins across the samples studied. The observed peptide intensities are approximated by the peak areas of the ions originating from the isobaric tags used to label different samples. Several factors can affect the observed peptide intensities, such as the expression level of the protein that generates the peptide, some peptide specific features relating to different efficiency in ionization and fragmentation, different amounts of samples loaded into different channels, differences in sensitivity to instrument detection, sample preparation and experimental variations (Luo and Zhao, 2012).

Several comparative iTRAQ based studies were conducted in different crops. Zhu et al., 2018 performed a comparative proteomics analysis using samples of non-embryogenic callus (NEC), embryogenic callus (EC) and somatic embryo (SE) using the isobaric tags for relative and absolute quantitation (iTRAQ) technology (Zhu et al., 2018). Another iTRAQ based study was conducted in *Nicotiana benthamiana* in which a leaf proteome identifies important proteins secondary metabolite biosynthesis and defense pathways crucial to cross protection against TMV (Das et al., 2019). iTRAQ based proteomic analysis was employed to study the differences in protein expression profiles in chlorophyll-deficient and normal green leaves in the tea plant cultivar “Huangjinya” (Dong et al., 2018). A comparative proteome analysis of seedling leaves of 2 genotypes differing in salinity tolerance at 200 mM (18.3 dS/m) NaCl stress in cotton (*Gossypium hirsutum* L.) through iTRAQ technology. A total of 58 differentially

expressed salt-responsive proteins were identified (Gong et al., 2017). Large number of proteins showing differential expression were identified in stems and leaves from two alfalfa genotypes through iTRAQ analysis. Eleven differentially enriched proteins were mapped onto the PPP in stem, including two upregulated proteins and nine downregulated proteins (Sun et al., 2020). In coconut to understand the cold stress tolerance a comparative iTRAQ proteome profiling was done in two varieties Hainan Tall, BenDi (BD) and Aromatic coconut, XiangShui (XS). This study identified several differentially expressed genes stress response, photosynthesis and respiration related DEPs increased abundance in two varieties.

Several iTRAQ based studies were conducted in phenylpropanoid biosynthesis pathway. A comparative iTRAQ-based comparative proteomic analysis in mature leaves of Ethiopian mustard (*Brassica carinata*) provides insights into the whole-proteomic profiles and the role of anthocyanin biosynthesis in leaf color diversity were studied (Wang et al., 2019). In another study, a comparative transcriptome and iTRAQ-based proteome study was done to investigate the expression profile of the purple-leaved line (BC-P01) in comparison with green-leaved line (BC-G01) in *Brassica carinata* to elucidate the mechanisms of anthocyanin accumulation in the purple-leaved line (Wang et al., 2021). iTRAQ-Based quantitative proteomics analysis of black rice grain development reveals metabolic pathways associated with anthocyanin biosynthesis. Approximately 928 proteins were detected, of which 230 were differentially expressed throughout 5 successive developmental stages, starting from 3 to 20 days after flowering (DAF) were identified more over ACN biosynthesis, sugar synthesis, and the regulation of gene expression were upregulated, particularly from the onset of black rice grain development and during development (Chen et al., 2016).

The recent availability of platform technologies for high throughput proteome analysis has led to platforms that integrate messenger RNA and protein expression data (Hack, 2004). Several studies have compared the transcriptome and proteome data to get more insights into functional characterization studies. In *Chrysanthemum* (*Chrysanthemum* × *morifolium*) proteome involved in light induced anthocyanin biosynthesis were identified through iTRAQ. In this study light-induced structural and regulatory genes involved in the anthocyanin biosynthetic pathway in *Chrysanthemum* were identified using RNA sequencing and validated using iTRAQ technology (Hong et al., 2019). A comparative transcriptome and proteome analysis was conducted to find

mechanism involved in yellowing in Chinese yam (*Dioscorea opposita* Tunb.). Proteins and genes were significantly annotated to the same pathways, which indicated that expression of genes and proteins were significantly correlated. Earlier studies have proven the strength of an integrated approach employing transcriptomics and proteomics for authentication of genes of biosynthetic pathways (Fan et al., 2011). Hence, a study was formulated with the objective to validate biosynthesis pathway genes of curcumin by proteomic approaches.

3.3. MATERIALS AND METHODS

3.3.1. Plant material:

Rhizomes were collected after 120 days of planting from high-curcumin turmeric- IISR Prathibha, low curcumin turmeric- Accession No. 200 and a closely related species that is practically devoid of curcumin- *C. aromatica* from the Experimental Farm, Chelavoor of ICAR-Indian Institute of Spices Research (ICAR-IISR), Kozhikode, Kerala, India (Figure 1)

(A) (B) (C)

Figure 1: Cross section of (A) IISR Prathibha, (B)Acc.200, (C) *C. aromatica*

3.3.2. Tissue Lysis and protein extraction:

Three plant tissue samples were retrieved from -80°C deep freezer and thawed at room temperature. Tissues were weighed and approximately 2 gm of tissues were crushed in liquid nitrogen using mortar and pestle. Once the fine powder of tissue formed, they were transferred to a 50 mL falcon tubes. Prior to protein extraction, samples were washed in 10% Trichloroacetic acid (TCA) in 100% chilled acetone for 1 hour at -20°C and 1 hour in chilled acetone alone at -20°C.

Method 1: Protein extraction by Urea cell lysis buffer

Table 1: Urea cell lysis buffer composition

Materials	Vendor	Stock	Amount (1mL)	Amount (30mL)
Urea (7M)	Sigma-Aldrich	N/AP	0.42 gm	12.6 gm
Thiourea (2M)	Sigma-Aldrich	N/AP	0.152 gm	4.56 gm
SDS (2% v/v)	Sigma-Aldrich	20% Stock	100µL	3000µL
PMSF (2mM)	Roche	0.1M Stock	20µL	600µL
Sodium orthovanadate	Sigma-Aldrich	100 mM	10µL	300µL
Sodium pyrophosphate	Sigma-Aldrich	125mM	20µL	600µL
β-glycerophosphate	Sigma-Aldrich	1 M	1µL	30µL

The samples were centrifuged at 10,000 g for 12 min, supernatant was discarded and pellets taken for cell lysis and protein extraction. All samples were mixed with 5 mL of lysis buffer (Table 1) and incubated at 4°C for 30 min. Later, the samples centrifuged at 10,000 g for 15 min at 4°C. The supernatant was transferred to chilled acetone for overnight protein precipitation.

The samples were centrifuged at 9,000 g for 30 min at supernatant were removed and pellet were washed with 80% acetone. Then the pellets were dissolved in 5 mL of 50 mM Triethylammonium bicarbonate (TEABC) buffer.

Method 2: Protein extraction by Phenol cell lysis buffer

Table 2: Phenol cell lysis buffer composition

Materials	Vendor	Concentration
HEPES (pH 7.5)	Sigma-Aldrich	50mM
Sucrose	Sigma-Aldrich	40%
EDTA-Na	Sigma-Aldrich	1mM
Sodium orthovanadate	Sigma-Aldrich	100 mM
Sodium pyrophosphate	Sigma-Aldrich	125mM
β-glycerophosphate	Sigma-Aldrich	1 M

Acetone washed sample were subjected to cell lysis and protein extraction using Table 2 extraction buffer. Samples were incubated at 4°C for 10 min on rocking. Then, equal amount of Tris-HCL saturated phenol was added to each sample and mixed completely and kept at 4°C for 20 min on rocking. Later, samples were centrifuged at 3,000 g for 15 min and the supernatant was taken for acetone precipitation as mentioned in Method 1.

3.3.3. SDS-PAGE and protein digestion

Protein extract were subjected for SDS-PAGE to measure the amount of protein and quality check. We used 30 µL of each sample and run at 10% SDS-PAGE for 2 hours at 80V and 120V. The gel was separated and stained using Colloidal Coomassie blue solution for 15 min, followed by destaining using 50:40:10 of Methanol: Water: Acetic acid. To reduce the cysteine bridges, samples were incubated with 10 mM of DTT for 45min at 56°C. Post reduction, to stop rearrangement of cysteine bridges, the samples were alkylated with 20 mM of IAA for 30 min under dark at RT. All six samples, post reduction and alkylation were checked for their pH and optimized to pH 8 with 50 mM TEABC and water. Trypsin enzyme was added to the samples at 1:20 ratio (Enzyme: Protein) and kept for overnight incubation at 37°C. Digested samples were subjected to SDS-PAGE. Twenty-five microliters of digested samples were taken and run 10% SDS-PAGE for 2 hrs at 80V and 120V. The gel was separated and stained using Colloidal coomassie blue solution for 15 min, followed by destaining using 50:40:10 of Methanol: Water: Acetic acid.

3.3.4. Desalting by Waters SepPak C18 1cc Cartridge:

The peptides formed are desalted to remove all the DTT, IAA and Urea present in the samples by Waters SepPak C18 1cc cartridge. One sample with 100µg equivalent peptides were diluted further 10 times using 0.1% FA in water (0.1% FA). Desalting was initiated with activation of C18 bed by passing through 100% acetonitrile, followed by conditioning with 0.1% FA twice. Later, peptide mixture samples were loaded and passed through the C18 cartridge slowly, hence all the peptides will be available for the C18 bind. Washing of C18 cartridges post sample loading were performed twice using 0.1% FA to remove DTT, IAA and Urea. Now, peptides bound to the C18 cartridge were eluted slowly by passing 50% ACN in water solution twice.

3.3.5. Desalting by StageTip C18 Cartridge:

In order to check the efficiency of C18 cartridges, rhizome sample was taken and 25µg equivalent peptide sample was re-suspended using 100 µL of 0.1% FA. C18 StageTips were prepared in advance, 3M™ Empore™ C18 Extraction Disks were used. The StageTips prepared using C18 disks were initially activated by passing through 100% ACN and conditioned with 0.1% FA. Later, peptide mixture samples were loaded and passed through the C18 StageTips slowly, hence all the peptides will be available for the C18 binding. Washing of C18 StageTips post sample loading were performed twice using 0.1% FA. Now, peptides bound to the C18 StageTips were eluted slowly by passing 50% acetonitrile in water solution twice. The complete desalting step using C18 StageTips were performed with the help of MicroCL 21R Microcentrifuge at the speed of 500 g under RT.

3.3.6. LC-MS/MS data acquisition:

All the three desalted samples were dried and dissolved using 100µL and 25µL of 0.1% FA for C18 cartridge and StageTip sample, respectively. The same were loaded on to 96 well plate and placed in the sample manager of Easy-nLC-1200. Around ~1 µg of peptide mixture was loaded on to the Acclaim PepMap™ 100 trap column (75µm X 2 cm, nanoViper, C18, 3µm, 100Å) at a flow rate of 4 µL/min. Peptides were then made to pass through and separate using PepMap™ RSLC C18 (2µm, 100Å, 50µm × 15 cm) analytical column. The peptides bound to C18 stationary phase of the analytical column were eluted out using two mobile phases A *i.e.* 0.1% FA in water and B *i.e.* 0.1% FA in 80% ACN in gradient mode for 30 min. Below is the gradient method used for the peptide elution.

The online elution of peptides out of PepMap™ RSLC C18 analytical column were ionized and detected using Orbitrap Fusion Tribrid mass spectrometer. ESI technique was employed to ionize the peptides under positive ionization mode using EASY-Spray™ Source. The data was acquired in DDA mode, where positively charged ions (Peptide precursors) ranging between 400-1600 m/z were filtered using quadrupole and trapped in C-trap until AGC reaches 2e5 for 10 ms or either of the first, followed by ion detection in Orbitrap under the resolution of 120 K. Ions were then isolated further using quadrupole with 2Da window based on their intensity and charge state, trapped in C-trap until AGC reaches 1e5 for 200 ms or either of the first, followed by fragmentation using HCD technique. A 35±3% of NCE was applied and the fragments ranging between 110-2000 were acquired at 60K.

3.3.7. Data analysis and Database search and validation of identified proteins using qRT-PCR:

Identification of the spectra and the DEPs was performed by Redcliffe Life Sciences Pvt. Ltd. All three raw files were searched for peptides and proteins separately using Mascot and Sequest HT search engines present in Proteome Discoverer 2.2. The proteomic database search was performed against Araport *Arabidopsis* proteome database (48,359 protein sequences). Protein sequences were theoretically digested using protease trypsin with maximum 2 missed cleavages and searched against the LC-MS/MS generated peptide precursor and fragment ions (b and y ions) with 10 ppm and 0.02 Da peptide and fragment mass tolerance respectively. Modifications such as, carbamidomethylation at cysteine (C), oxidation at (M) and acetylation (Protein N-termini) were included as fixed and variable modifications. To calculate the FDR, decoy database was generated by reversing the amino acid sequences at peptide and protein level. The PSM's were identified with FDR less than 1% (q-value <0.01) are considered as true matches.

Gene Ontology (GO) functional enrichment analysis of the identified proteins was conducted using Blast2GO (Conesa et al., 2005). Annotation of proteins was identified using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) before KEGG pathway enrichment analysis by KOBAS (Xie et al., 2011).Correlation between the proteomic and the transcriptomic libraries was analyzed using Microsoft Excel based on proteomic data obtained in the present study and transcriptomic data (Bio-project ID PRJNA698442). Primers were designed for identified proteins using primer quest tool and validated using qRT-PCR.

3.4. RESULTS AND DISCUSSION

Proteomics technology has been widely used and has yielded great advances in scientific research (Aslam et al., 2017). iTRAQ is one of the most flexible techniques that allow multiplexing of four, six, and eight separately labelled samples within one experiment, with a reduction of costly LC–MS runtime (Dowle et al., 2016; Elliott et al., 2009). In this study, iTRAQ-based technology was used to validate the DEG's identified through transcriptomics and expression analysis by analysing their proteins whose abundances also differed significantly between high and low curcumin accession.

3.4.1. Protein estimation and digestion

The reliability and efficiency of protein identification and quantitation in iTRAQ is dependent on sample preparation intended to reduce complexity upstream of MS (Spanos and Moore, 2016). Extraction of proteins from cells using chemical buffers is challenging as it is required to ensure that the proteins remain in-solution for subsequent digestion. Among the two methods we used for protein isolation, extraction by urea cell lysis buffer yielded more protein than phenol cell lysis buffer. The details are provided in Figure 2 and Table 3. Though urea-based buffers are widely used in proteomics (Guo et al., 2007), some of the studies reveal that extraction buffer with SDS yield a greater number of proteins (Dapic et al., 2017). It is also advisable to remove excess urea before the addition of trypsin, which may interfere with the subsequent digestion step (Brownridge and Beynon, 2011)

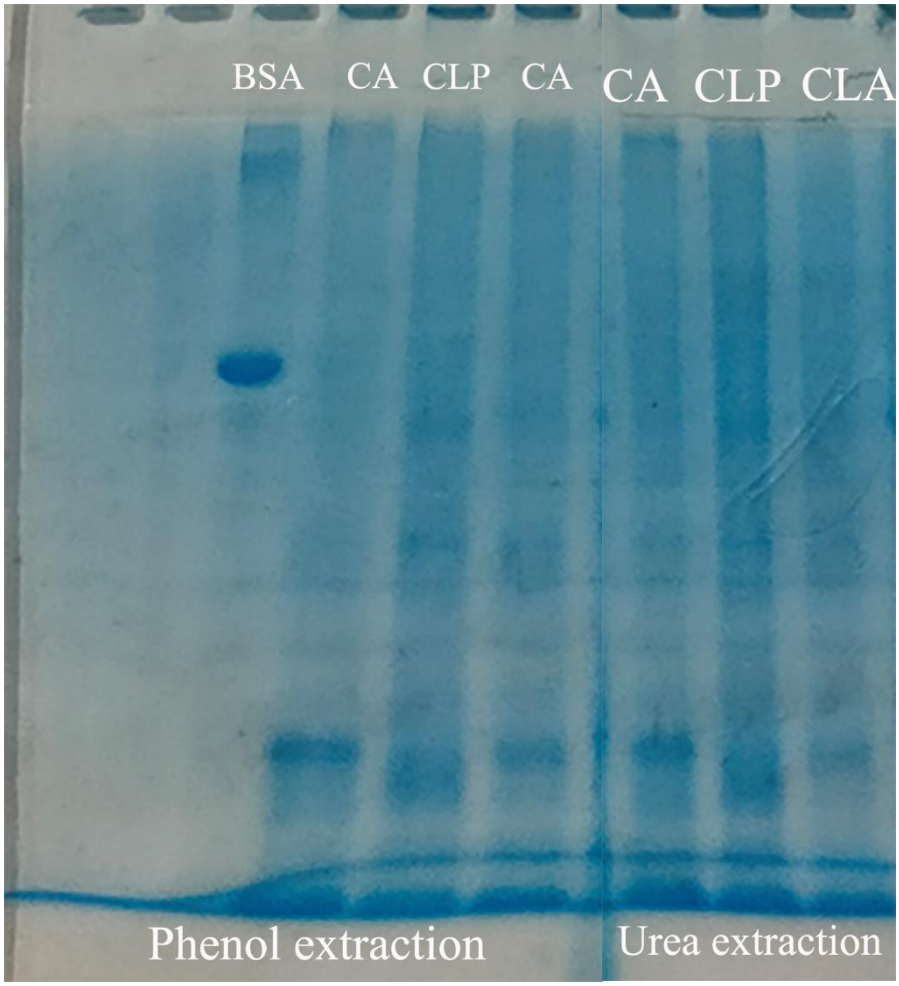


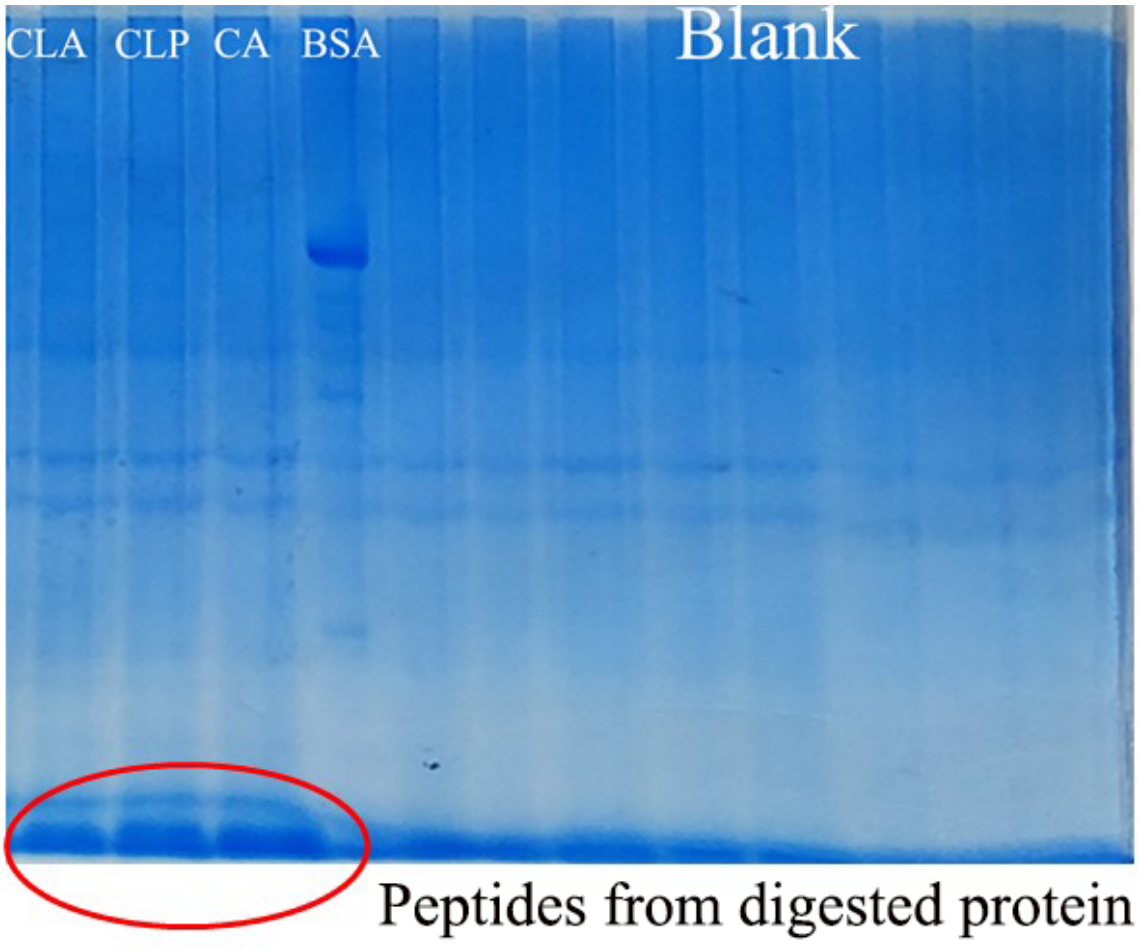
Figure 2: Protein estimation by SDS-PAGE * 5µg BSA was loaded as control

Table 3: Protein amount estimated for Urea and Phenol lysis buffer extraction methods based on SDS-PAGE

Sample Details	Load volume (µL)	Protein amount from Urea extraction (µg/30µL)	Protein amount from Phenol extraction (µg/30µL)	Total protein yield from Urea extraction (mg)
CLP	30	6	5	1.4
CLA	30	10	7	2.34
CA	30	6	4	1.4

* The protein extraction was done from 2mg of tissue

The digestion of proteins by trypsin has resulted in peptides (Figure 3)



* 5µg BSA was loaded as control

Figure 3: Post-Digestion check by SDS-PAGE

3.4.2. Proteome search result:

Raw data files from all the three samples (CLP, CLA and CA) were used for search against *Arabidopsis* proteome database, *Zingiberales* Proteome database and *Curcuma* six-frame translated database and around 100, 1627 and 3058 proteins could be identified respectively for each sample. The details of the result from different database search types are provided in Table 4.

Table 4: Summary of raw reads searched against different databases

Search Type	Database	MS/MS spectra	PSMs	Peptides	Proteins
Zingiberales Proteome Search	Zingiberales (Uniprot)	285591	38620	4032	1627
Curcuma six-frame translated database search	*Curcuma* six-frame (In-house)	239644	47192	6543	3058
Arabidopsis Proteome search	*Arabidopsis* (Araprot-11)	214322	208	110	99

3.4.3. Overview of quantitative proteome analysis

In total, 214,322 unique spectra were identified from *Curcuma longa* using iTRAQ technique based on the three proteomic libraries of CLP, CLA and CA through *Arabidopsis* (Araprot-11) database search. 285, 591 and 239,644 unique spectra were identified though Zingiberales Proteome and *Curcuma* six-frame translated database search. Similarly, 302,042 unique spectra of differentially accumulated proteins were identified from chlorotic and green leaves of *Camellia sinensis* (Dong et al., 2018). Analysis using the Mascot software revealed that 110 unique peptides matched to available spectra. Subsequently, 99 raw proteins were assembled. Quality control of the protein data showed that the protein masses were distributed normally based on their expression abundance (Figure 4), indicating the high quality of identified proteins in the present study, suitable for further analyses. Molecular mass of the identified protein ranged from 10 kDa to 200 kDa and calculated PI ranged between 4.0 to 13 (Figure 5).

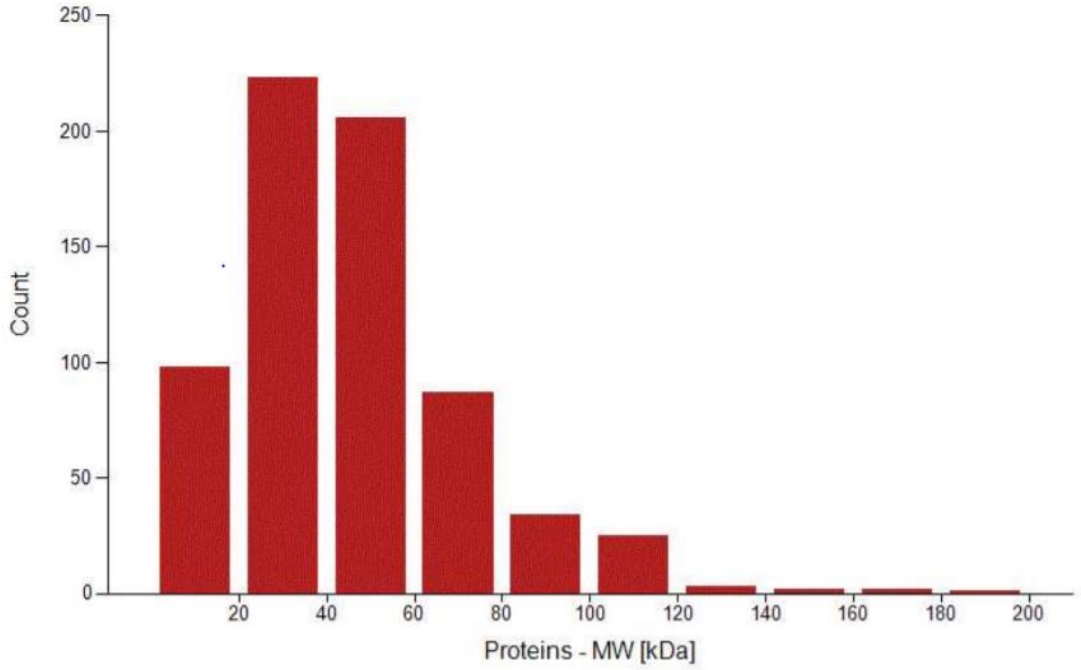


Figure 4: Histogram showing protein molecular weight distribution

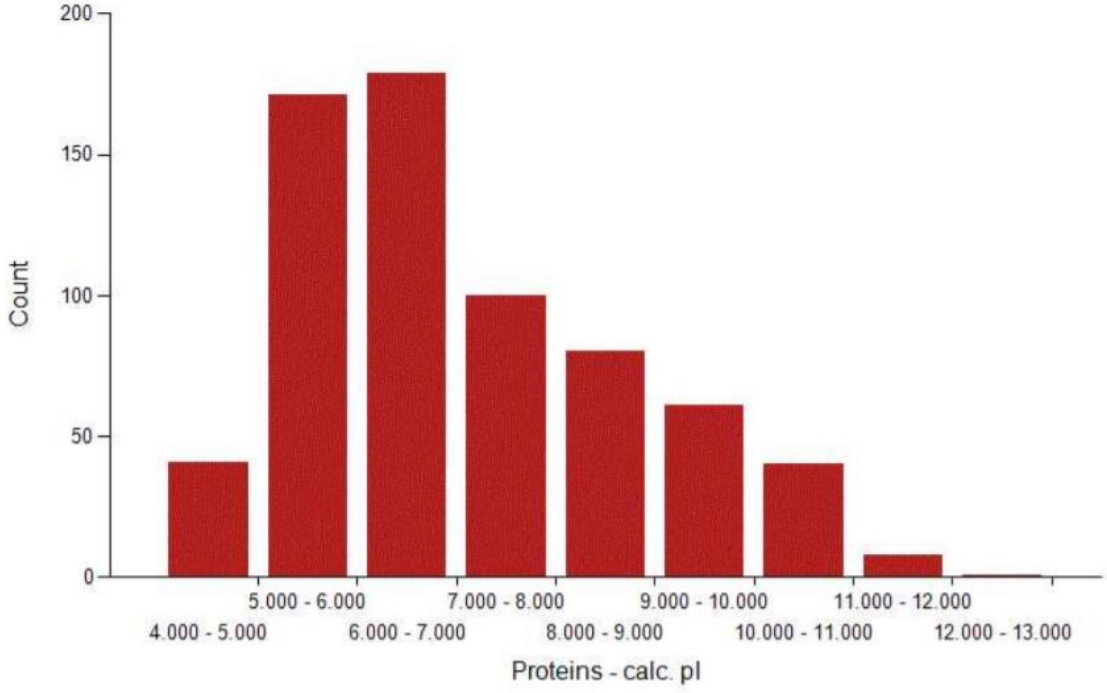


Figure 5: Histogram showing protein isoelectric point distribution

3.4.4. Correlation between the protein expression and the curcumin biosynthesis under different curcumin condition

Sample intersection with proteomics libraries was done involving CLP vs. CLA, CLP vs. CA and CLA vs. CA where CLP in order to investigate the correlation between protein expression and curcumin biosynthesis (Figure 6). For identification of differentially expressed proteins cut off points were fixed as P-value<0.05 and fold change>=1.5. Volcano maps were constructed, where red is used to represent data points showing a differential expression. The vertical coordinate was log10 (p-value); the smaller the p-value, the more significant the difference, while the farther the ordinate value is from 0, the bigger the difference. Each point in the volcano map represents a protein. The red dots in the upper left corner and the upper right corner represent the differentially expressed proteins, and the black spots are non-differential proteins (Figure 6). A total of 1567 proteins were identified with high confidence, with 23 differentially expressed proteins identified related to curcumin biosynthesis pathway including phenylalanine ammonia lyase, caffeoyl CoA O-methyltransferase, curcumin synthase 1 and diketide CoA synthase and transcription factors such as WD40, NAC, bZIP *etc*.

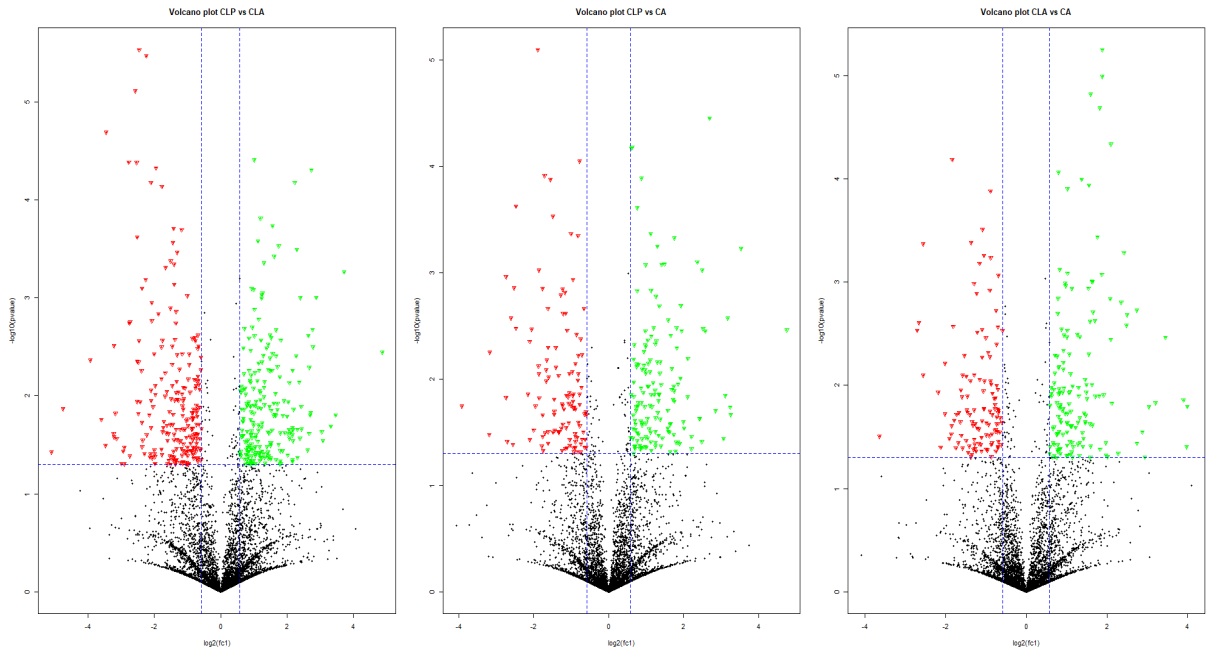


Figure 6: Volcano plot representing differentially expressed peptides in CLP vs. CLA, CLP vs. CA and CLA vs CA; red color indicating upregulated proteins and green color representing downregulated proteins where as black color represents nonsignificant proteins.

3.4.5. Hierarchical Clustering Analysis

Hierarchical clustering analysis identified that genes involved in curcumin biosynthesis such as NAC, C4H, WD40, Homeobox domain-containing protein, ABC transporter proteins, HTH transcription factor *etc*. that showed positive correlation and expressed high abundance in IISR Prathibha and low expression in Acc 200 and *C. aromatica*. (Figure 7).

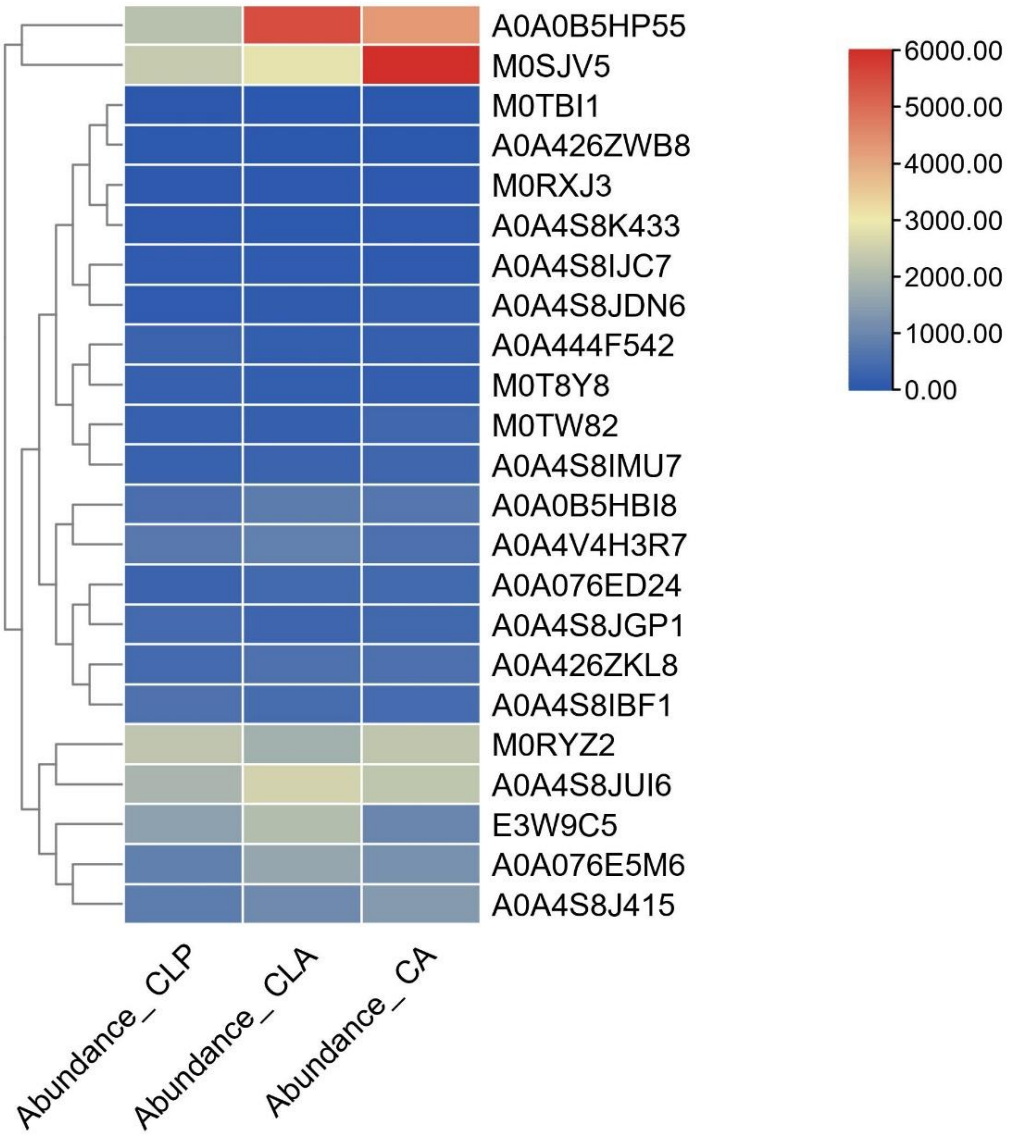


Figure 7: Hierarchical clustering analysis of curcumin biosynthesis pathway proteins obtained from iTRAQ using CLP, CLA and CA

3.4.6. Functional annotation of differentially expressed proteins

3.4.6.1. Molecular function

The GO, KEGG and COG annotation of the identified proteins were carried out to comprehensively reflect the biological function and significance of these proteins in various life activities in which molecular function including translation regulator activity (GO:0045182), molecular transducer activity (GO:0060089), binding (GO:0005488), structural molecule activity (GO:0005198), molecular function regulator (GO:0098772) catalytic activity (GO:0003824), and transporter activity (GO:0005215) were observed. Among them 47% of proteins had catalytic function (246 proteins), 35% were binding proteins (180 proteins) (Figure 8).

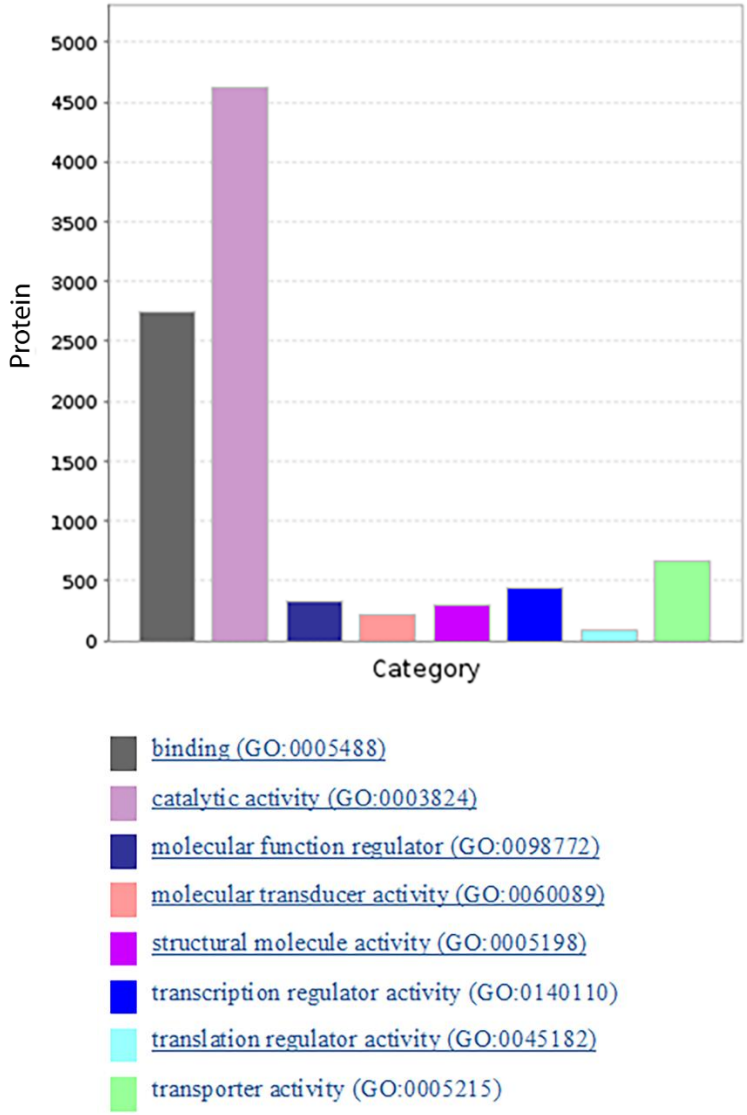
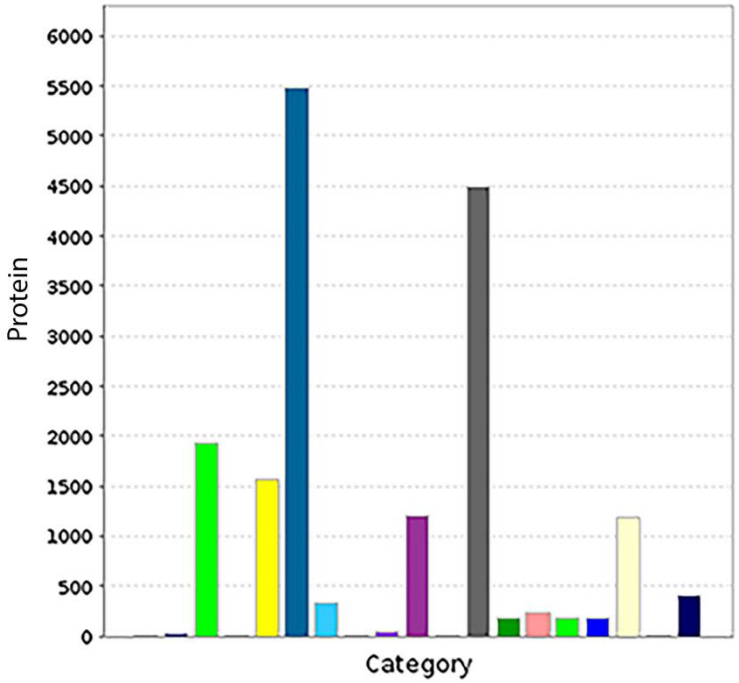


Figure 8: Different molecular functions of differentially expressed proteins

3.4.6.2. Biological processes

Biological processes including cellular component organization or biogenesis (GO:0071840), cellular process (GO:0009987), biological phase (GO:0044848), reproductive process (GO:0022414), localization (GO:0051179), reproduction (GO:0000003), biological regulation (GO:0065007), response to stimulus (GO:0050896), signalling (GO:0023052), developmental process (GO:0032502), multicellular organismal process (GO:0032501), and metabolic process (GO:0008152). In biological process 37% proteins were included in the cellular process followed by metabolic process. (Figure 9)



- biological phase (GO:0044848)
- biological regulation (GO:0065007)
- cell population proliferation (GO:0008283)
- cellular component organization or biogenesis (GO:0071840)
- cellular process (GO:0009987)
- developmental process (GO:0032502)
- growth (GO:0040007)
- immune system process (GO:0002376)
- localization (GO:0051179)
- locomotion (GO:0040011)
- metabolic process (GO:0008152)
- multi-organism process (GO:0051704)
- multicellular organismal process (GO:0032501)
- reproduction (GO:0000003)
- reproductive process (GO:0022414)
- response to stimulus (GO:0050896)
- rhythmic process (GO:0048511)
- signaling (GO:0023052)

Figure 9: Different biological process of differentially expressed proteins

3.4.6.3. Cellular component

Cellular component including plasmodesma (GO:0009506), membrane part (GO:0044425), membrane (GO:0016020), organelle part (GO:0044422), extracellular region part (GO:0044421), cell junction (GO:0030054), membrane-enclosed lumen (GO:0031974), protein-containing complex (GO:0032991), supramolecular complex (GO:0099080), extracellular region (GO:0005576), cell (GO:0005623), cell part (GO:0044464), and organelle (GO:0043226). 54% of these proteins are in cell or part of the cell (Figure 10).

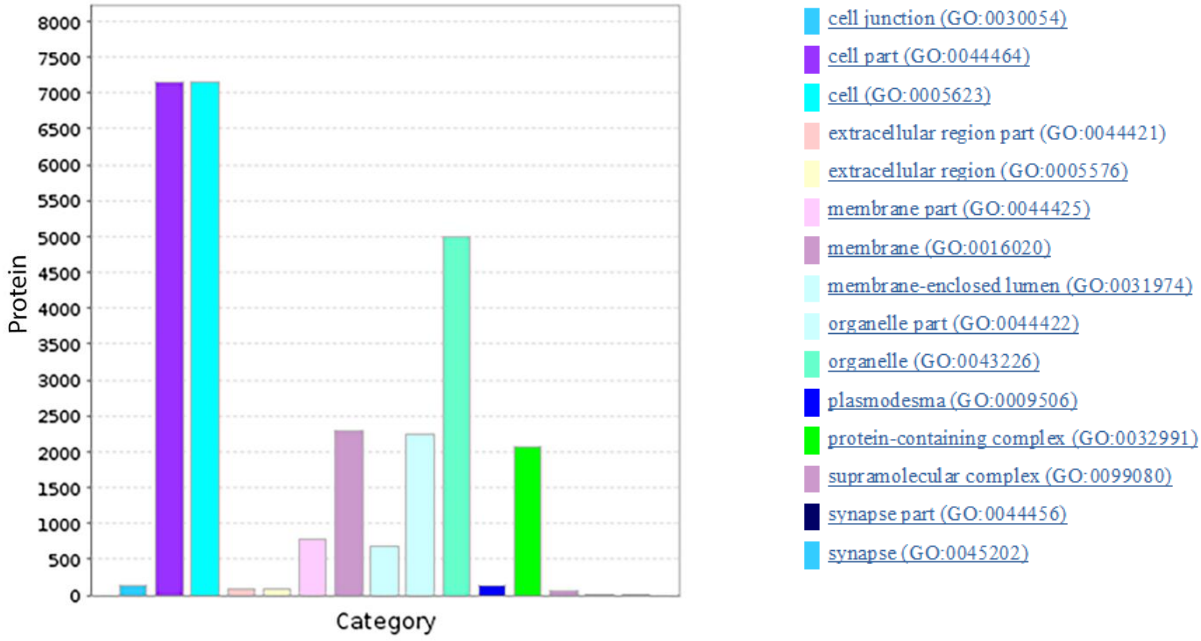


Figure 10: Cellular component analysis of differentially expressed genes

3.4.7. Pathway analysis

Protein motifs includes protein modifying enzyme (PC00260), transporter (PC00227), scaffold/adaptor protein (PC00226), membrane traffic protein (PC00150), chaperone (PC00072), protein-binding activity modulator (PC00095), transfer/carrier protein (PC00219), nucleic acid binding protein (PC00171), transmembrane signal receptor (PC00197), defense/immunity protein (PC00090), calcium-binding protein (PC00060), cytoskeletal protein (PC00085), translational protein (PC00263), metabolite interconversion enzyme (PC00262), chromatin/chromatin-binding, or -regulatory protein (PC00077). Majority of the proteins are metabolite interconverting enzymes (Figure 11).

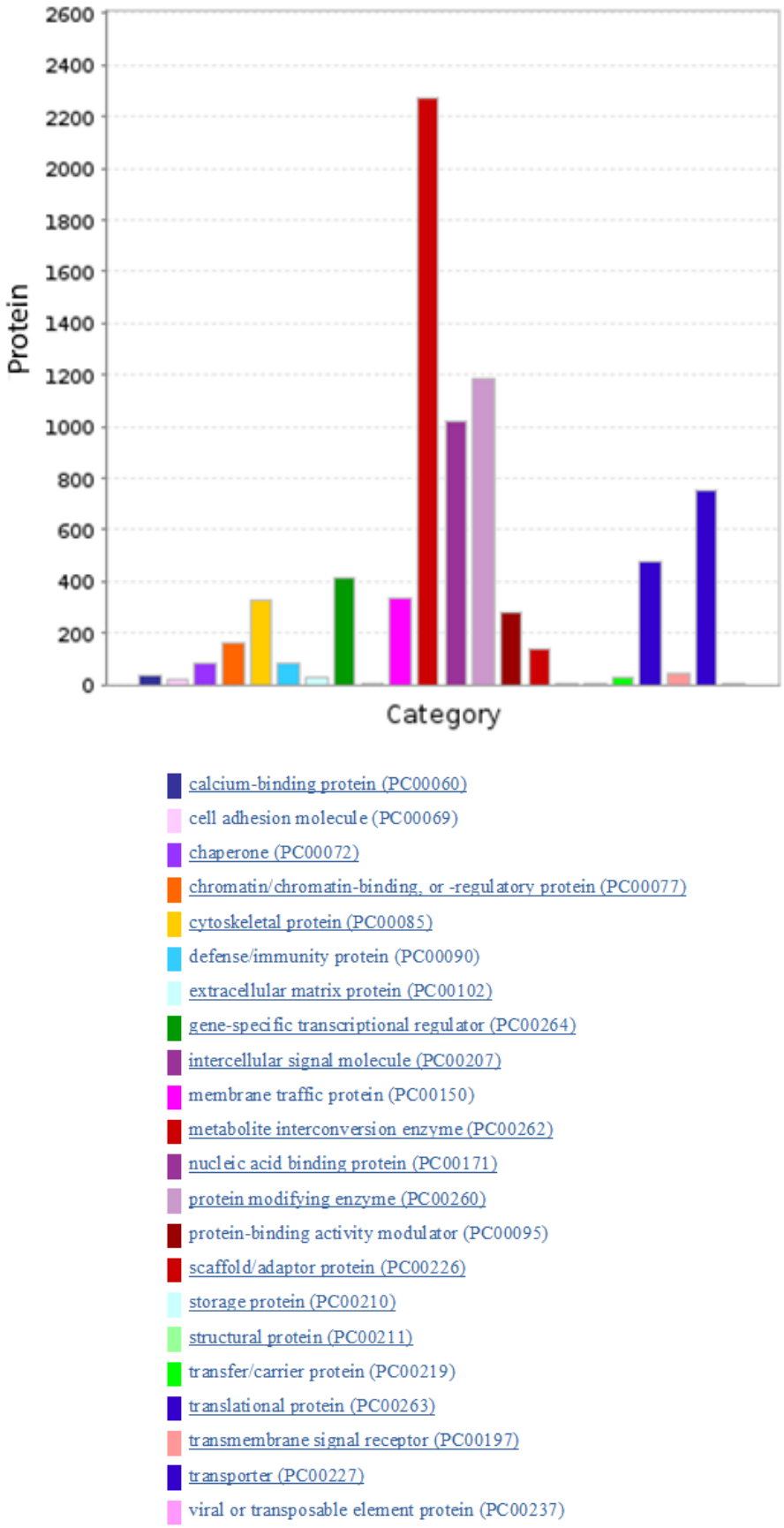


Figure 11: Differentially expressed genes and the respective protein class

3.4.8. Identification of curcumin biosynthesis pathway genes and transcription factors from iTRAQ proteome data

Several transcription factors such as WD40, F box protein, HTH, NAC domain containing proteins were identified through proteome data. Moreover, curcumin biosynthesis pathway genes such as phenylalanine ammonia lyase, cinnamate-4-hydroxylase, caffeoyl CoA O-methyltransferase, curcumin synthase 1, diketide CoA synthase and two polyketide synthases were also identified. Phenylalanine ammonia lyase identified from proteome is identical in sequence with PAL identified from transcriptome (Figure 12). Sequence of caffeoyl CoA showed identity with already reported Caffeoyl CoA O-methyltransferase (Figure 13). CURS1(Figure 14) and DCS (Figure 15) showed high identity with amino acid sequence obtained from proteome data. Rest of the proteins such as C4H, ClPKS11 and WD40 sequences did not show perfect identity with the transcriptome derived sequences.

Table 5: Significantly differentially expressed proteins in phenylpropanoid biosynthetic pathway

Acession no.	Description	Molecular weight	Abundance
E3W9C5	p450 mono-oxygenase (Cinnamate 4 hydroxylase) OS=*Zingiber zerumbet* OX=311405 GN=CYP71BB1 PE=2 SV=1	505	1540
A0A076E5M6	Coffee acyl coenzyme A-3-O-methyl transferase OS=*Curcuma longa* OX=136217 PE=2 SV=1	245	873.1
A0A0B5HP55	Curcumin synthase 1 OS=*Curcuma longa* OX=136217 GN=CURS1 PE=2 SV=1	389	2173.8
A0A0B5HBI8	Diketide-CoA synthase 1 OS=*Curcuma longa* OX=136217 GN=DCS1 PE=2 SV=1	389	498.9
A0A076ED24	Phenylalanine ammonia lyase (Fragment) OS=*Curcuma longa* OX=136217 GN=PAL PE=2 SV=1	389	283.8
A0A4S8IJC7	PKS_ER domain-containing protein OS=*Musa balbisiana*OX=52838GN=C4D60 _ Mb09t16410 PE = 4 SV=1	348	119.5
A0A4S8K433	PKS_ER domain-containing protein OS=*Musa balbisiana*OX=52838GN=C4D60_ Mb08t15740 PE=3 SV=1	372	78.2

Table 6: Significantly differentially expressed transcription factor proteins identified from proteome data

Accession no.	Description	Molecular weight	Abundance
M0RYZ2	NAC-A/B domain-containing protein OS=*Musa acuminata subsp. malaccensis* OX=214687 GN= 103988636 PE=4 SV=1	200	2277.3
A0A426ZKL8	WD_REPEATS_REGION domain-containing protein OS=*Ensete ventricosum* OX=4639 GN=B296_ 00041818 PE=3 SV=1	302	451
A0A4S8IBF1	WD_REPEATS_REGION domain-containing protein OS=*Musa balbisiana* OX=52838 GN= C4D60_Mb02t17610 PE=4 SV=1	342	574
M0TW82	HTH cro/C1-type domain-containing protein OS=*Musa acuminata subsp. malaccensis* OX=214687 GN= 103996559 PE=4 SV=1	207	142
M0RXJ3	NAC-A/B domain-containing protein OS=*Musa acuminata subsp. Malaccensis* OX=214687 GN= 103980412 PE=4 SV=1	214	74
A0A4V4H3R7	F-box domain-containing protein OS=*Musa balbisiana* OX=52838 GN=C4D60_Mb06t08290 PE=4 SV=1	396	726

```
A0A076ED24      SGGRNLSLDYGFKGAEIAMAAYCSEPQFLGNPVTNHVQSAEQHNQDVNSLGLISARKTAE
ClPAL2          SGGRNPSLDYGFKGAEIAMAAYCSELQFLGNPVTTHVQSAEQHNQDVNSLGLISARKTAE
                ***** ****************** ********.************************

A0A076ED24      AVDILKLMSATYLVALCQ-IDLRHLEENLKHAVKAAVSLVGKRVLTTGANGELHPARFCE
ClPAL2          AVDILKLMSATYLVALCQAIDLRHLEENLKQAVKTAVSLAAKRALTTGANGELHPARFCE
                ***************** ***********:***:****..**.***************

A0A076ED24      KDLLTVVDREHVLSYADDPCSSAYVLMPKLRMVLVEHALGHGDKEKDAAASIFHKVAAFE
ClPAL2          KDLLTVVDREHVLSYADDPCSSTYLLMPKLRMVLVEHALGHGDKEKDAAASIFHKIAAFE
                **********************:*:********************************:****

A0A076ED24      EELKAVLPKEVEAARAALESGNPAISIRIKECRSYPLYRLVREQLGAAYLTGEKVRSPGE
ClPAL2          EELKAVVPKEVEAARAAVENDSAAIGNRIKECRSYPLYRLVREDLGAAYLTGEKVRSPGE
                ******:**********:*....**. ****************:****************

A0A076ED24      EFEKVSAAINAGLVIDPLLQCLNEWNGAPIPIC--------------------
ClPAL2          EFEKVFAAINAGLLVDPLLDCLKEWNGAPIPSAKPPGPPAKKKKKKKKSTLR
                ***** *******::****:**:********* .
```

Figure 12: Clustal w analysis of ClPAL2 identified from transcriptome with caffeoyl CoA O-methyltransferase identified from iTRAQ

```
A0A076E5M6      MCVTHQEKHNRLISPKKTKTQREKEREMATTTTEATKTSSTNGEDQKQSQNLRHQEVGHK
ClOMT3          ----------------------------------------MASDSKNGEQHRHQEVGHK
                                                        ..*.*:.:: ********

A0A076E5M6      SLLQSDDLYQYILETSVYPREPESMKELREVTAKHPWNIMTTSADEGQFLNMLIKLVNAK
ClOMT3          SLLQSDALYQYILETSVYPREPEAMKELRDITAKHPWNLMTTSADEGQFLGMLLKLINAK
                ****** ****************:*****::*******:***********.**:**:***

A0A076E5M6      NTMEIGVYTGYSLLATALALPEDGKILAMDVNRENYELGLPIIEKAGVAHKIDFREGPAL
ClOMT3          NTMEIGVYTGYSLLATALALPDDGKILAMDINRENYEIGLPVIQKAGVAHKIDFREGPAL
                **********************:********:******:***:*:****************

A0A076E5M6      PVLDEIVADEKNHGTYDFIFVDADKDNYINYHKRLIDLVKIGGVIGYDNTLWNGSVVAPP
ClOMT3          PVLDQLLEDAKNHGSFDFVFVDADKDNYINYHKRLLELVKVGGVIAYDNTLWNGSVVAPP
                ****::: * ****::**:*****************::***:****.**************

A0A076E5M6      DAPMRKYVRYYRDFVLELNKALAADPRIEICMLPVGDGITICRRIS
ClOMT3          DAPMRKYIRYYRDFVLELNKALAADERIEICQLPVGDGVTICRRVK
                *******:****************** ***** ******:*****:.
```

Figure 13: Clustal W analysis of ClOMT3 identified from transcriptome with curcumin synthase identified from iTRAQ

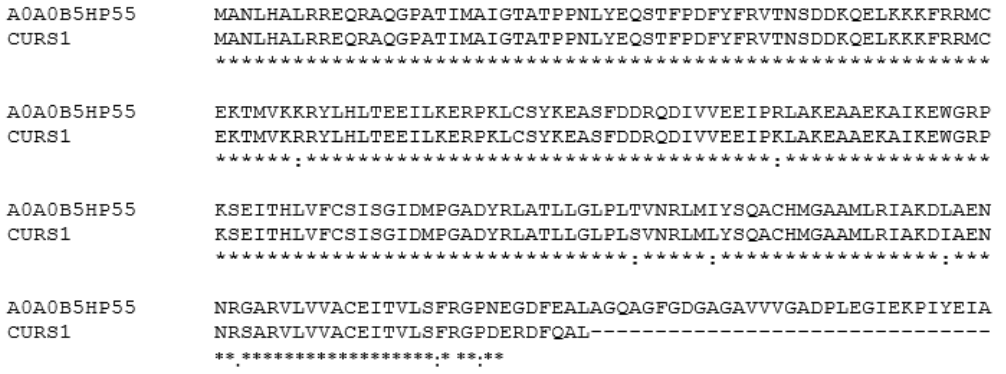


Figure 14: Clustal W analysis of CURS1 identified from transcriptome with diketide CoA synthase identified through iTRAQ

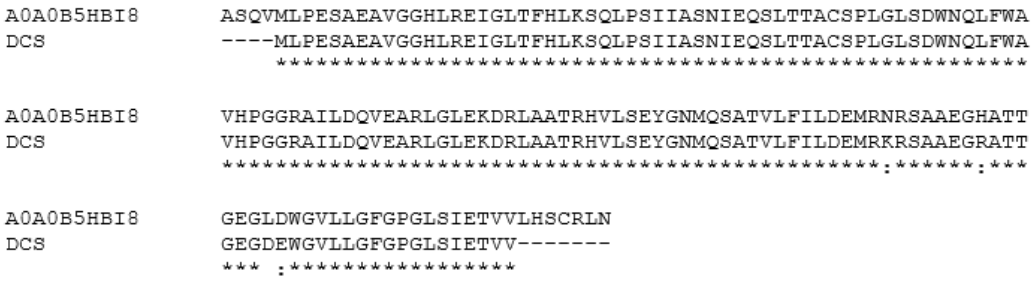


Figure 15: Clustal W analysis of DCS identified from transcriptome with diketide CoA synthase identified through iTRAQ

3.4.9. Validation of iTRAQ data by qRT-PCR

To better understand the molecular mechanisms involved in biosynthesis pathway, a combined analysis of the transcriptome and proteome are essential (Dong et al., 2019). In this study, we combined an iTRAQ-based proteome-level analysis with an RNA sequencing-based transcriptome-level analysis to detect the proteins and genes related to curcumin biosynthesis in *Curcuma longa*. We analyzed the IISR Prathibha, Acc. 200 and *C. aromatica* cultivars at 4th month of development. In this study to further validate the proteome data, qRT-PCR was performed based on the sequences obtained from the iTRAQ and transcriptional patterns of phenylalanine ammonia lyase, caffeoyl CoA O-methyltransferase, diketide CoA synthase and curcumin synthase were evaluated. EF1α was used as the reference gene. Interestingly, the transcript levels of these four genes were significantly upregulated (fold change -51.98,315.1,103.9, and 34.29 for

PAL, CCoAOMT, CURS1, and DCS respectively). Relative expression levels (FPKM) for these genes were 10.61, 37.67, 76.85, and 39.3 respectively. Relative abundances of the proteins in iTRAQ were 283.8, 873.1, 2173, and 498.9 for PAL, CCoAOMT, CURS1, and DCS respectively. These results indicated a correlation between transcriptome data and proteome data (Figure 16). A similar study conducted in sweet orange infected by ‘‘*Candidatus Liberibacter asiaticus’* also showed a correlation between iTRAQ proteome and transcriptome data of the defense related genes and proteins (Fan et al., 2011).

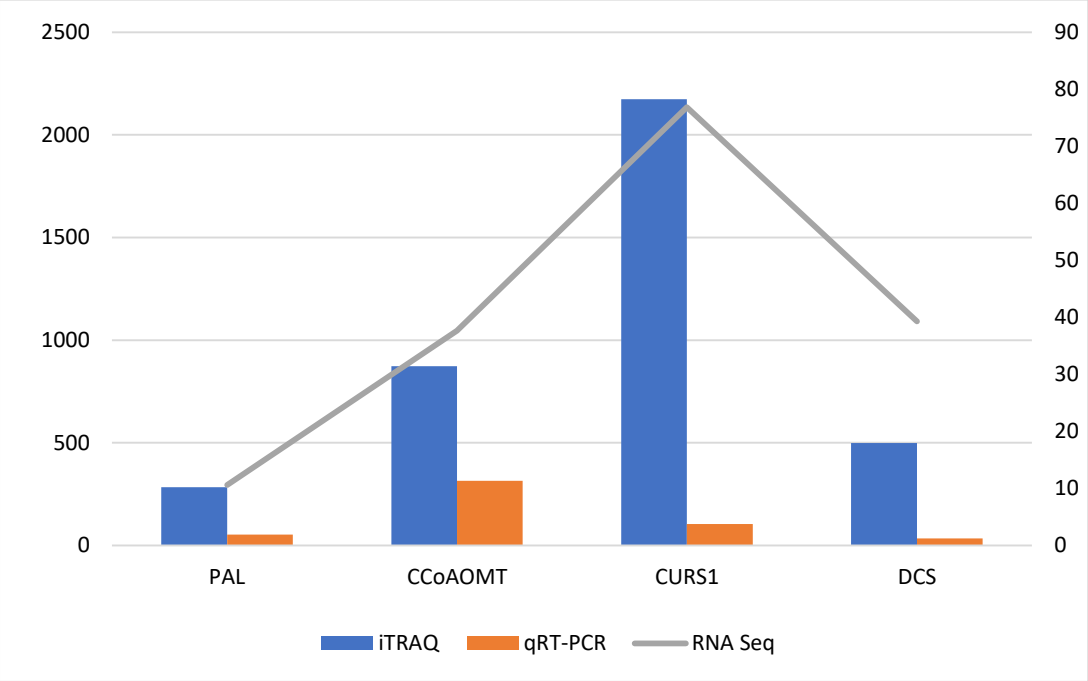


Figure 16: validation of iTRAQ data by qRT-PCR

3.5 CONCLUSION

The iTRAQ mediated analysis of whole protein profiles helped in identifying proteins such as phenylalanine ammonia lyase, caffeoyl CoA O-methyltransferase, diketide CoA synthase and curcumin synthase 1 differentially expressed from rhizomes of turmeric. The study indicated an obvious difference in profiles of the differentially expressed proteins among the high and low curcumin lines. These differentially expressed proteins, correlated well with the corresponding gene counterparts identified from a comparative transcriptome based on contrasting curcumin. These proteins and genes can be considered as biomarkers to indicate high curcumin levels. Being dedicated candidates of curcumin biosynthesis they may be used individually or combined to indicate the curcumin status of the crop. Further works aimed at deciphering the functions of these proteins may add greater insights into their roles in biosynthesis.

Chapter-4

qRT-PCR BASED BIOMARKER FOR CURCUMIN IN TURMERIC (*Curcuma longa* L.)

4.1. INTRODUCTION AND OBJECTIVES

High curcumin varieties are preferred by pharmaceutical industries and curcumin industry; however, curcumin content is found to vary with growth and management conditions and varieties used (Guddadarangavvanahally et al., 2002; Anandaraj et al., 2014). The major bottleneck faced by curcumin industry is the lack of easy and fast screening techniques for curcumin during early stages of development. Conventional methods of curcumin estimation involve elaborate sample preparation and estimation procedures and require minimum 1-week time and the sample is usually retrieved after the crop attains full maturity at 8 months from planting. The HPLC based procedures also is elaborate and time consuming and more costly than conventional methods. All these procedures require destructive harvesting of the plant. Turmeric is a vegetatively propagated crop and selection for high curcumin lines in breeding programs is done under field condition entirely based on curcumin content analyzed by conventional methods after harvest. This is time consuming, laborious and requires field trials over 8 months duration. By contrast, MAS or marker assisted selection strategies if available can make this screening process cheaper, faster and may be less prone to errors caused by environment and growth conditions and can be implemented during early stages of development. Curcumin being the most important trait, effective screening strategies for selection of high curcumin lines during the early stages of development may also help breeders in fast screening of populations. Morphological markers for curcumin are not available and DNA/RNA markers would be of much advantage due to their robustness, ease of detection, ability to screen large populations, independence to genotype, environment, development stage and time. Molecular marker like transcripts and metabolites, are advantageous because they integrate factors like environmental effects and many genes. Curcumin being a secondary metabolite is a complex phenotype and actively synthesized in tandem with the development stages of the plant and DNA/RNA based markers would be better. However, as we know turmeric is a polyploid and vegetatively propagated crop and use of DNA markers may be problematic. Gene expression-based biomarker would be an ideal strategy in case of curcumin. Several gene expression markers based on transcriptome data is available in crops like maize,

sunflower *etc*. for assessing various plant responses and status (Rengel et al., 2012; Marchand et al., 2013). We developed a qRT-PCR based simple screening strategy for screening high/low curcumin lines during early stages of development at 4^{th} month that is before crop attains maturity, by a non-destructive method. However, we can expect several challenges like sample reproducibility and immediate arresting of metabolic activities at the time of sampling *etc*. as it is an RNA based method. Considering the huge amount of RNA-seq data available with us as well as several genes and transcripts identified to have a potential role in curcumin biosynthesis, it is worthwhile to exploit the Next Gen sequencing platforms and Real-time PCR technologies to develop such novel strategies. These marker candidates have been tested in independent sets of samples including blind samples and the robustness and applicability was confirmed.

We have used RNA-seq data from high and low curcumin turmeric accessions as well as related species of *Curcuma* that is low in curcumin. For confirming the biomarkers, we have used different types of tissues like rhizomes and leaves from the above samples. Blind samples were provided by a collegue from his breeding population. Three candidate genes of the biosynthetic pathway for curcumin *viz*., ClPKS11, ClPAL2, ClOMT2 and ClOMT3 were chosen for evaluating their suitability as biomarker for curcumin. These genes were chosen based on our earlier experiments where the expression profiles showed a positive correlation with curcumin. These genes are respectively homologous to already reported genes from other crops All these genes were validated under field conditions as well as greenhouse conditions and showed a positive correlation with curcumin irrespective of environment, variety, management and growth conditions as mentioned in Chapter 1. The objective of the study was to demonstrate the possibility to characterize the curcumin status from expression levels of key candidate genes and to develop a practical tool that could be easily used for determining curcumin status in a non-destructive manner.

4.2. REVIEW OF LITERATURE

A biomarker is defined as a measurable and/or observable change in a biological or biochemical response, ranging from the molecular to the physiological level and including behavioural changes (Joshi et al., 2015). Diagnostic biomarkers have been discovered in medicine many decades ago and are now commonly applied. Several studies were conducted to identify molecular markers associated with human diseases

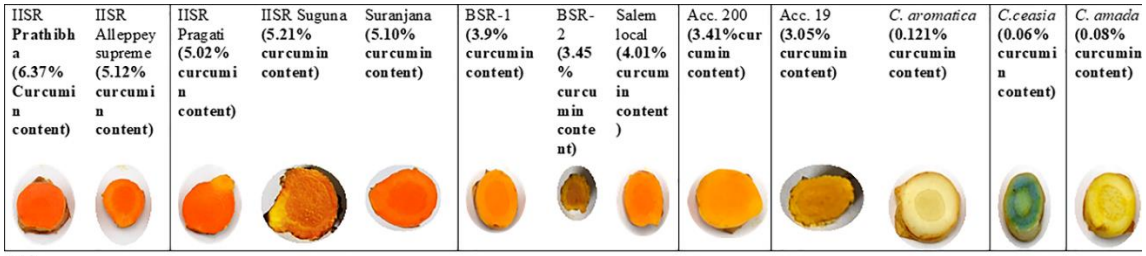
(Amato et al., 2017; Chen et al., 2019, 2018). While this is routine in the field of medicine, in plants biomolecular based study was less. In plants biomarkers are used to predict phenotypical properties before these features become apparent and, therefore, are valuable tools for both fundamental and applied research. (Steinfath et al., 2010) discovered plant metabolic biomarkers for phenotype prediction using an untargeted approach and application of metabolomics to predict agronomic important phenotypes of a crop plant that was grown in different environments. By using this combination of metabolomics and biomathematical tools metabolites were identified that can be used as biomarkers to improve the prediction of traits. Another study was conducted *Avena sativa* to identify signatory biomarkers for cultivar identification using metabolomics was applied to unravel differential metabolic profiles of various oat cultivars (Magnifico, Dunnart, Pallinup, Overberg and SWK001) and (Pretorius et al., 2021). The primary step involved in plant breeding, inspection, registration, trade and seed production requires the identification of cultivars and varieties, and therefore a rapid and effective method for cultivar fingerprinting is required (Korir et al., 2013). Over the years, plant breeding has been greatly improved for unravelling the molecular basis of complex traits using genomic analyses and next-generation sequencing methods (Edwards et al., 2013). Currently, plant breeding methods have integrated phenotypic traits with a range of marker-assisted selection techniques to more efficiently determine trait outcomes (Gebhardt, 2013). Although genetic markers have been at the forefront of plant breeding efforts, many limitations have restricted the use thereof in analysing complexities arising from genotype × environment interactions (environmental plasticity), polygenic inheritance and epistasis (which is referred to as the action of one gene on another) (Yandeau-Nelson et al., 2015). A metabolite and transcript markers were developed in potato (*Solanum tuberosum* L) for the prediction of potato drought tolerance through qRT-PCR based study. In this study transcript marker candidates were selected from a published RNA-Seq data set. By combining transcript marker with metabolic marker to predict the drought tolerance (Sprenger et al., 2018). A study was conducted for the identification of drought tolerance markers in a diverse population of rice cultivars by expression and metabolite profiling. They identified 46 candidate genes with significant genotype X environment interaction in an expression profiling study on four cultivars with contrasting drought tolerance. Among those with significant positive correlation was the gene coding for a cytosolic fructose-1,6-bisphosphatase. This enzyme catalyses a highly regulated step in Carbon metabolism. The metabolic and transcript marker

candidates for drought tolerance were identified in a highly diverse population of cultivars. Thus, these markers may be used to select for tolerance in a wide range of rice germplasms (Degenkolbe et al., 2013).

4.3. MATERIALS AND METHODS

4.3.1. Plant material

To develop a biomarker for screening high and low curcumin lines we used five high curcumin (IISR Prathibha, IISR Alleppey Supreme, IISR Pragati IISR Suguna and Suranjana), five low curcumin (BSR-1, BSR-2, Salem local, Acc. 200, and Acc. 19) and three low curcumin related species (*C. aromatica*, *C. amada* and *C. ceasia*) collected from the Experimental Farm, Chelavoor of ICAR-Indian Institute of Spices Research (ICAR-IISR), Kozhikode, Kerala, India. Three different tissues such as rhizome sprouts, leaves and primary rhizomes of five high, low curcumin accessions and three low curcumin related species were used (120 DAP, 150 DAP and 180 DAP) (Figure 1).



(A)

Figure 1: (A) Five high curcumin accessions five low curcumin accessions and three low curcumin-related species of turmeric (*Curcuma. longa* L.). (B) Field view of five high five low and three very low curcumin relates species planted at Experimental farm, Chelavoor, ICAR-IISR, Kozhikode, Kerala state, India

4.3.2. Transcriptome mining, selection of biomarkers and expression analysis

Two transcriptomes generated under contrasting curcumin contents were used for selection of suitable biomarker candidates (BioProject ID PRJNA270561 and PRJNA277549 for *C. longa* and *C. aromatica*; PRJNA698442 for IISR Prathibha under field and IISR Prathibha under stress). We focussed on genes of higher expression in high curcumin background compared to low curcumin backgrounds, since practically what is required is to identify high curcumin lines from a population and high transcript levels are easier and more accurate to measure compared to low transcript levels. We have shortlisted four prominent genes that invariably showed a positive correlation with curcumin under different experimental conditions in the field *viz., CLPKS11, CLPAL2, CLOMT2, CLOMT3*. RNA isolation and cDNA synthesis was done as mentioned in Chapter 2. Four reference genes such as EF1α, ubiquitin, tubulin, GAPDH were tried for sprout, leaves and rhizomes. Stable reference gene were identified using RefFinder (RefFinder (heartcure.com.au)). Representation of stable reference gene was done using box plot generated using a web-based tool BoxplotR (shiny.chemgrid.org/boxplotr/). Real time PCR was performed as mentioned earlier in Chapter 2. Primers for real time was designed using primer quest tool (PrimerQuest Tool | IDT (idtdna.com) and listed in Table 1.

Table 1: qRT-PCR primers of selected contigs were designed using primer quest tool (PrimerQuest Tool | IDT (idtdna.com))

	Forward primer (5'-3')	**Reverse primer (5'-3')**
Quantitative real time PCR primers to amplify candidate reference genes		
EF1	CGGCTGCTCTGAGAAACAAT	GGTGATGCTGGACAAACAGTAG
UBIQUITIN	AAACCCAGTGGAGCAACTT	TATCGCTTGCGAGGCATATC
GAPDH	CAAGAGGAGCAAGACAGTTAGT	CAGATGCTCCTATGTTTGTGATTG
TUBULIN	GCGGAAGCAAATGTCGTAAAG	GGTAGAGCCATACAATGCTACTC
Quantitative real time PCR primers to biomarker		
ClPKS11	TGTCGGAGATCACCCACTTG	CGGAGAAGGGAGACCGAGA
ClPAL2	TTGGATTGTTGTTGGTGGTTTAG	GAGGAGTTCGAGAAGGTGTTC
ClOMT1	CTGCTCACCAAGCTCTCTATTC	TAGGCCTCTTCTTCGTCGT
ClOMT2	GCGTCTACACGGGCTATTC	CAGCCCAATCTCGTAGTTCTC

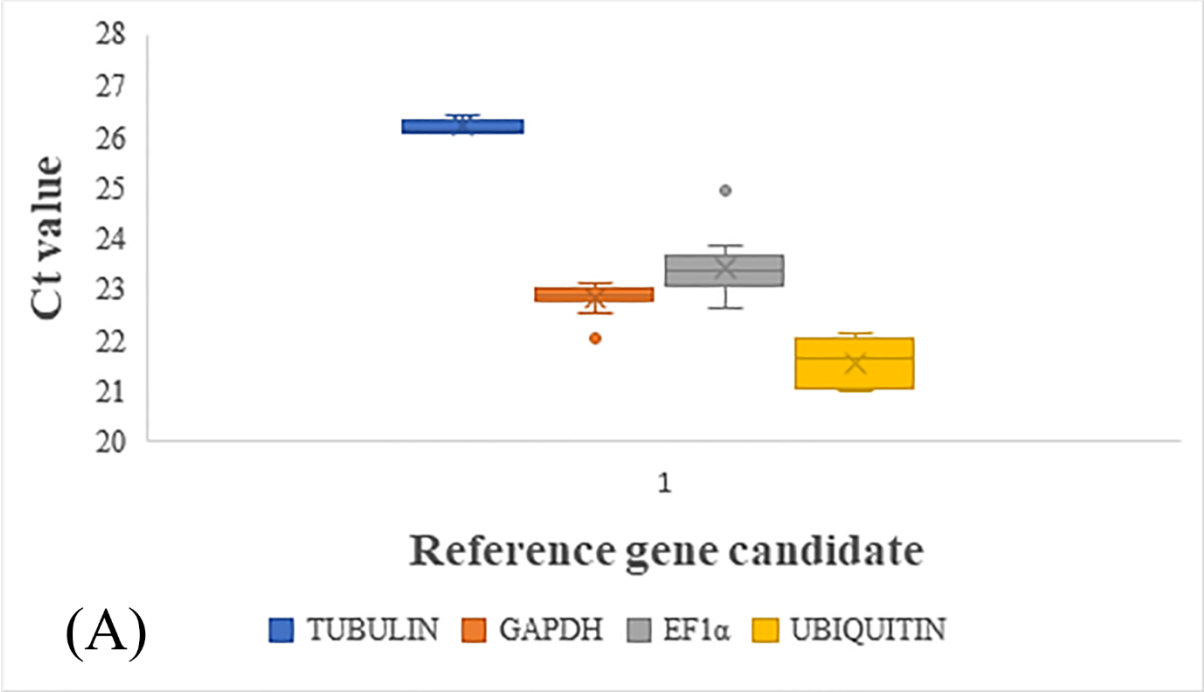
4.3.3. Transcriptome mining of biomarkers and expression analysis

Curcumin status is correlated with a change in expression of key candidate genes of the biosynthetic pathway for curcumin. This distinguishes high and low curcumin lines independent of variety, growth conditions, tissue used for analysis *etc*. So far, no reports on correlation of curcumin status with transcript level/gene expression is available. It needs to be noted that the method developed by us is different from conventional method of identifying high curcumin lines. The variation in curcumin status within the experimental samples taken is significant enough to use it as a basis for discovery of markers for curcumin.

Comparative transcriptome was used as the source for mining of biomarkers and housekeeping genes based on the fold change and FPKM values. Primarily we shortlisted important transcripts that gave a positive correlation with curcumin content from the transcriptome data (Table 2). Along with *CLPKS11* other important genes such as *ClPAL2*, *ClOMT2* and *ClOMT3* were also evaluated for their suitability as biomarker for curcumin. And role of these genes was confirmed using different environmental as well as experimental conditions. The curcumin status biomarker was defined from the expression levels of these four genes normalised by the expression levels of reference genes. Reference genes were validated under each biological context and the one showing less variability among samples was considered as the stable reference gene. Tubulin was identified as the stable reference for sprout tissue (Figure 2) and *EF1α* was identified as the stable reference gene for both leaves and rhizome (Figure 3 and 4).

Table 2: Transcriptome mining for the identification of biomarker and house keeping genes for the study

Contig ID	Sequence length (bp)	FPKM value in high curcumin condition	FPKM value in low curcumin condition	Sequence description
691222 (ClPKS11	326	695.55	10.04	Type III polyketide synthase 2
c45063_g3_i2 (cClPAL2)	961	10.61	1.24	Phenylalanine ammonia lyase, partial [*Curcuma longa*]
c42730_g1_i2 (ClOMT2)	1121	108.08	3.55	trans-resveratrol di-O-methyltransferase-like [*Elaeis guineensis*]
C41020_g1_i2 (ClOMT3)	1124	37.67	1.23	coffee acyl coenzyme A-3-O-methyl transferase [*Curcuma longa*]
Housekeeping genes/Reference genes				
659998	256			Elongation factor
599390	182			Ubiquitin-conjugating enzyme
737869	612			Glyceraldehyde-3-phosphate dehydrogenase
566568	158			Beta-tubulin

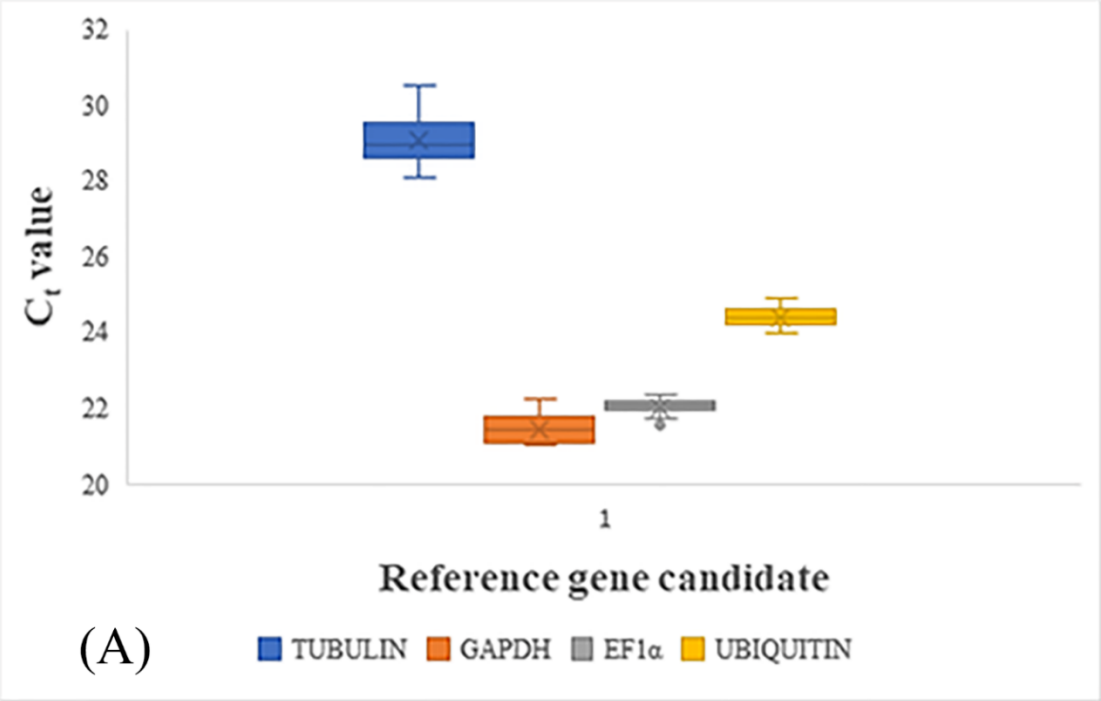


(A)

Ranking Order (Better--Good--Average)				
Method	1	2	3	4
Delta CT	TUBULIN	GAPDH	UBIQUITIN	EF1α
BestKeeper	TUBULIN	GAPDH	EF1α	UBIQUITIN
Normfinder	TUBULIN	GAPDH	UBIQUITIN	EF1α
Genorm	TUBULIN | GAPDH		UBIQUITIN	EF1α
Recommended comprehensive ranking	**TUBULIN**	**GAPDH**	**UBIQUITIN**	**EF1α**

(B)

Figure 2: (A) Expression of four candidate genes measured as Ct value by qRT-PCR in sprout tissue (B) ranking order of stable reference gene predicted using RefFinder

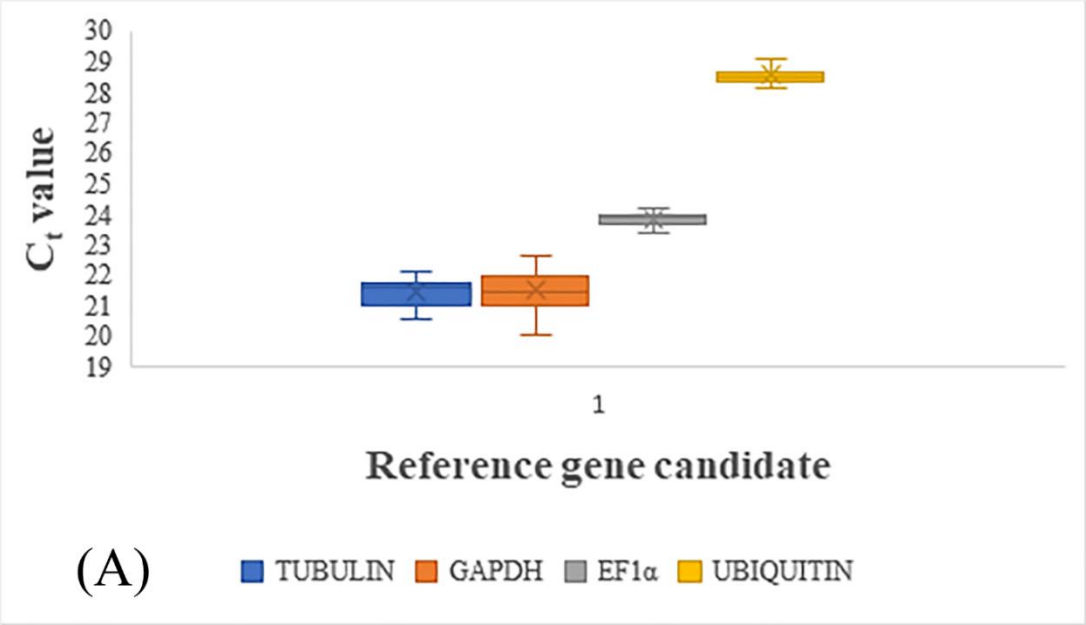


(A)

Ranking Order (Better--Good--Average)				
Method	1	2	3	4
Delta CT	EF1α	GAPDH	UBIQUITIN	TUBULIN
BestKeeper	EF1α	UBIQUITIN	GAPDH	TUBULIN
Normfinder	EF1α	GAPDH	UBIQUITIN	TUBULIN
Genorm	EF1α | UBIQUITIN		GAPDH	TUBULIN
Recommended comprehensive ranking	**EF1α**	**UBIQUITIN**	**GAPDH**	**TUBULIN**

(B)

Figure 3: (A) Expression of four candidate genes measured as Ct value by qRT-PCR in leaf (B) ranking order of stable reference gene predicted using RefFinder



Ranking Order (Better--Good--Average)				
Method	1	2	3	4
Delta CT	EF1α	UBIQUITIN	TUBULIN	GAPDH
BestKeeper	EF1α	UBIQUITIN	TUBULIN	GAPDH
Normfinder	EF1α	UBIQUITIN	TUBULIN	GAPDH
Genorm	EF1α | UBIQUITIN		TUBULIN	GAPDH
Recommended comprehensive ranking	**EF1α**	**UBIQUITIN**	**TUBULIN**	**GAPDH**

(B)

Figure 4: (A) Expression of four candidate genes measured as Ct value by qRT-PCR in rhizome (B) ranking order of stable reference gene predicted using RefFinder

4.3.4. Selection of marker transcripts

For discovery and validation of a robust marker for curcumin status, expression patterns of four genes such as *ClPKS11*, *ClPAL2*, *ClOM2*, and *ClOMT3* was verified both in the RNA-Seq data as well as through qRT-PCR. The first gene *ClPKS11* codes for a polyketide synthase identified from *curcuma* transcriptome through comparative transcriptome analysis. Total of sixty-three transcripts of polyketide synthases were identified from the RNA-seq data. Among them *ClPKS11* showed high FPKM (695.5) in high curcumin condition compared to FPKM (10.6) in low curcumin condition. The expression of *ClPKS11* correlated with curcumin content under different experimental conditions and *ClPKS11* showed maximum identity of 72% with already reported

CURS3 (curcumin synthase 3) and exhibited amino acid differences in the substrate binding pocket, cyclization pocket and geometry shapers surrounding the active site.

The second gene *ClPAL2* encoding phenylalanine ammonia lyase was also identified from RNA-seq data. Seven PAL transcripts were identified each with minor differences in the sequence. Among those transcripts, *ClPAL2* showed high FPKM (10.6) compared to that under low curcumin conditions (1.9). The expression of *ClPAL2* correlated with curcumin content under different experimental and environmental conditions. *ClPAL2* is a novel gene that showed 74.25% identity with *Hedychium coronarium.* Docking analysis showed that *ClPAL2* binds with phenylalanine at a low binding energy confirming its preference to the substrate. The third gene *ClOMT2* encodes O-methyltransferase and thirty O-methyltransferases were identified from RNA-seq data. Among these, *ClOMT2* showed high FPKM (108) in high curcumin condition and low FPKM (3.5) under low curcumin condition. Expression of *ClOMT2* correlated with curcumin in all the experimental and environmental studies. *ClOMT2* showed 90% similarity with trans-resveratrol O-methyltransferase-like protein from *Zingiber officinale*. Docking studies revealed that *ClOMT2* is involved in the downstream curcumin biosynthesis pathway and final conversion of curcumin.

The fourth gene *ClOMT3* encodes caffeoyl CoA O-methyltransferase and three caffeoyl CoA O-methyltransferases were identified from transcriptome. Among those transcripts, *ClOMT3* showed maximum FPKM value of 37.6 and 1.2 under low curcumin condition. *ClOMT3* showed 99.18% similarity with already reported coffee acyl CoA-3-O-methyl transferase from *Curcuma longa.* Docking analysis performed using Caffeic acid as substrate with lowest binding energy confirmed the substrate preference of *ClOMT3*.

Sampling materials were retrieved from five high curcumin and five low curcumin accessions and three very low curcumin related species involving sprouts from stored seed rhizomes, leaves and rhizomes at 120, 150 and 180 DAP. Sprouts were evaluated to verify if the biomarkers can be used at the very early stages for screening and leaves were chosen for the investigation as it is easy to obtain samples without disturbing the plant.

The expression of all the four genes was very low in all the sprout samples and correlated with the low curcumin status (Figure 5), however, in leaf samples, the

differential gene expression of four genes among low and high curcumin accessions was not observed, indicating lack of correlation between curcumin status and gene expression (Figure 6,7,8,9). However, in case of rhizomes, expression of *ClPKS11*, *ClPAL2*, *ClOMT2* and *ClOMT3* were upregulated in all the samples and expression correlated with curcumin status (Figure 10,11,12, and 13). Expression of *ClPKS11* was high in all high curcumin accessions and expression was maximum at 120 DAP and reduced at 150 DAP and further reduced at 180 DAP. Among rhizomes, leaves and sprout tissues taken for the study, rhizomes were found to be ideal for screening and among the three different stages (120 DAP, 150 DAP and 180 DAP) of development 120 DAP was found to be optimum. It is observed that Ct value of all high curcumin samples were below 22 whereas Ct values of all low curcumin samples were above 25.

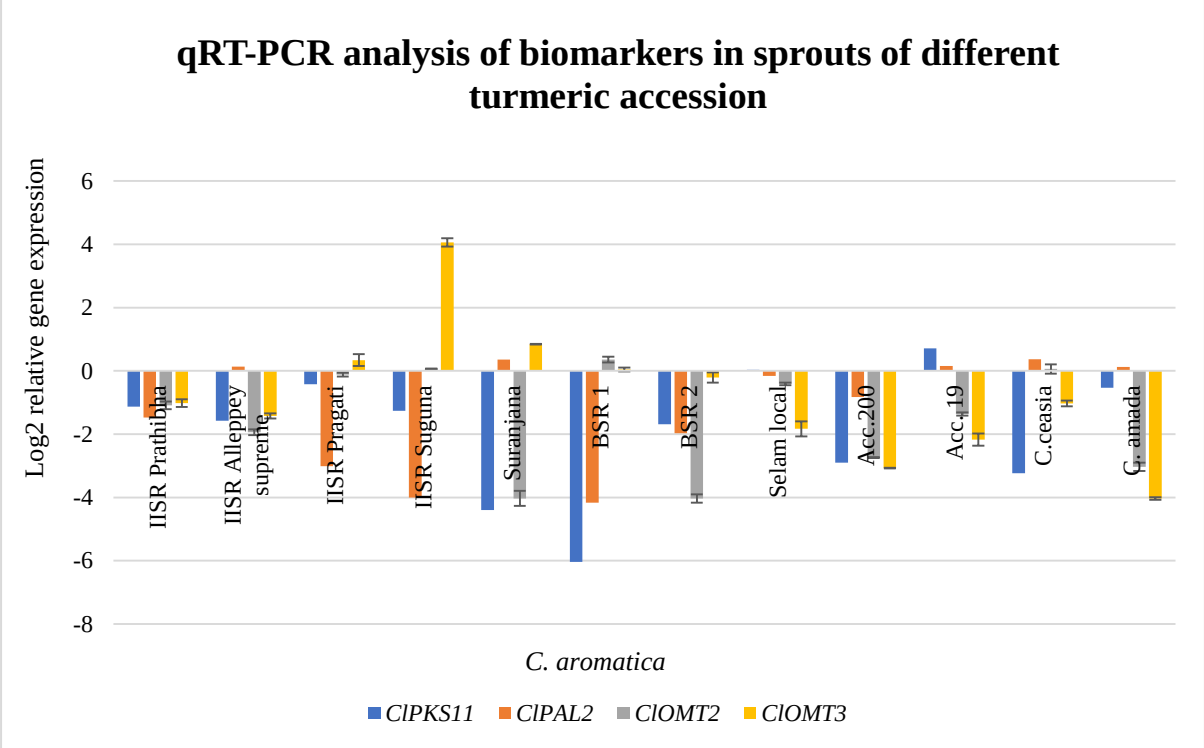


Figure 5: qRT-PCR analysis of *ClPKS11*, *ClPAL2*, *ClOMT2* and *ClOMT3*

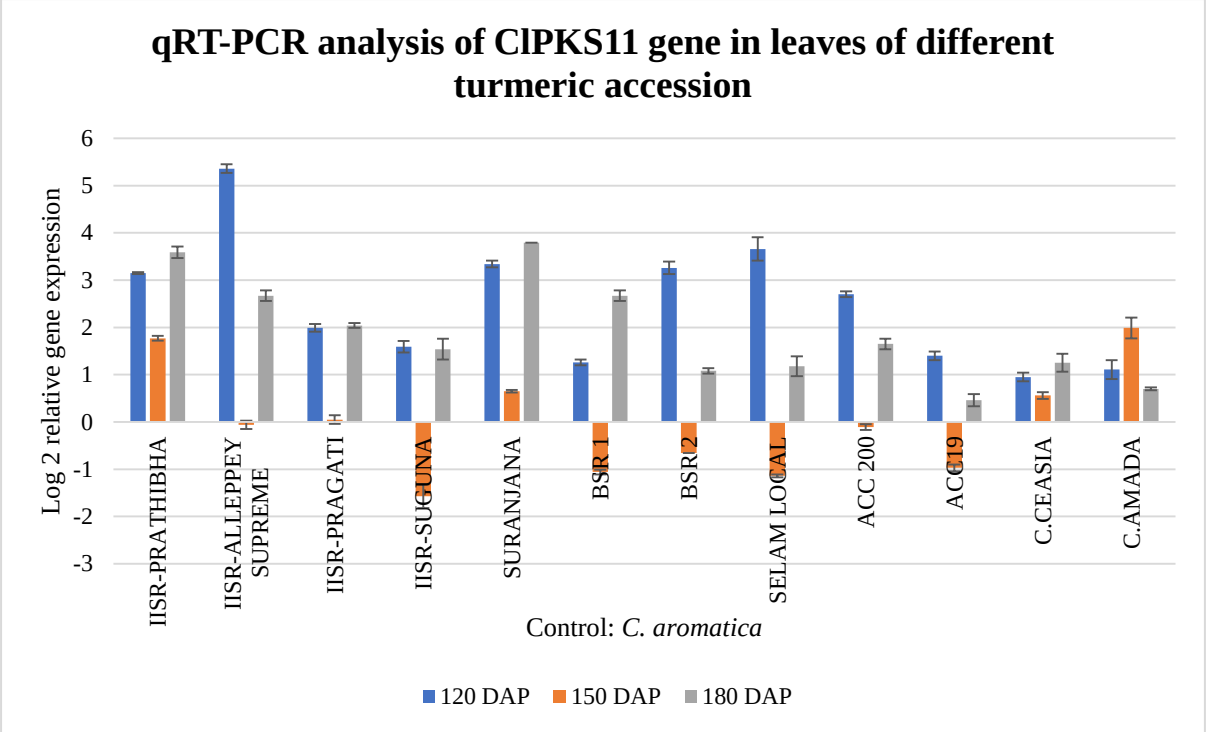


Figure 6: qRT-PCR analysis of *ClPKS11*under different developmental stages of leaves

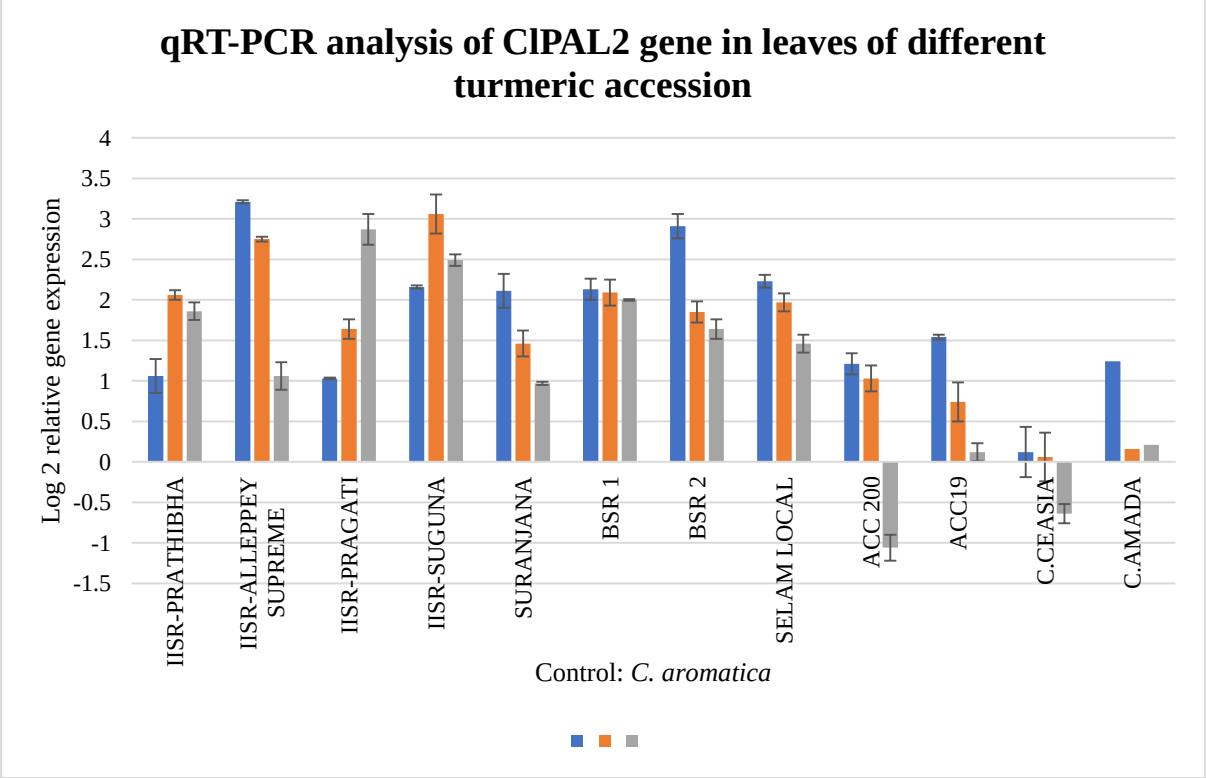


Figure 7: qRT-PCR analysis of *ClPAL2* under different developmental stages of leaves

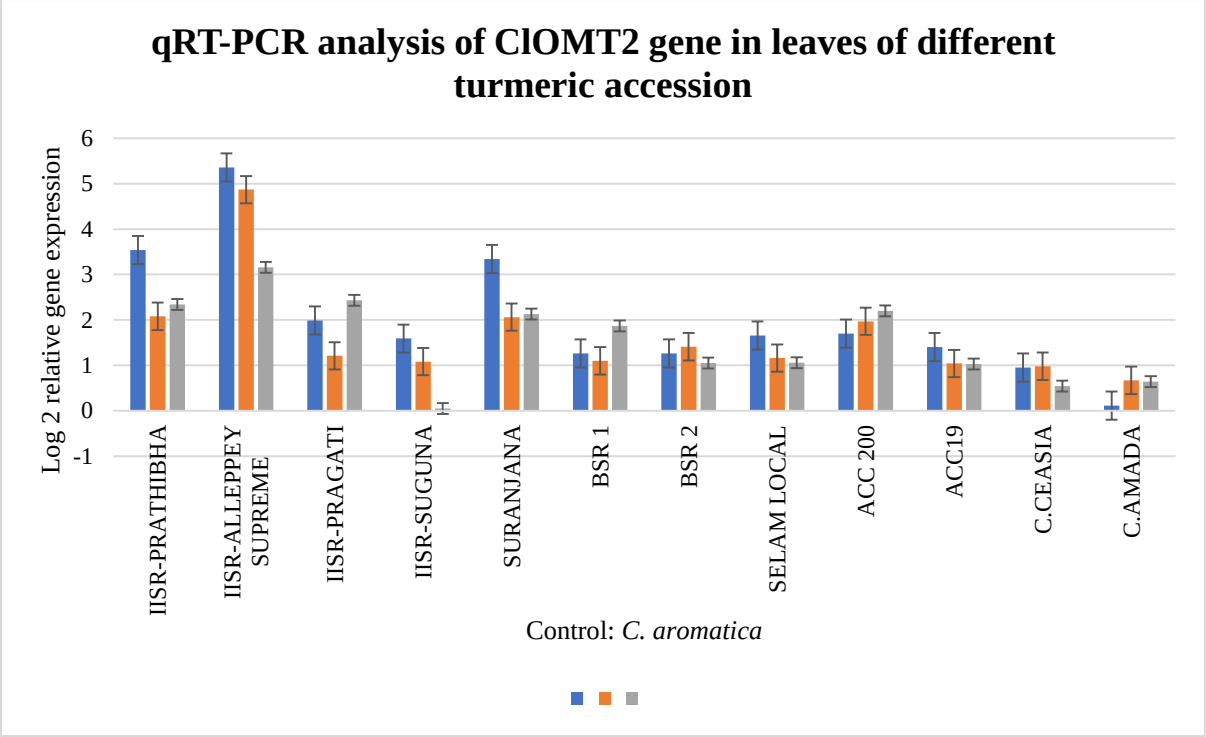


Figure 8: qRT-PCR analysis of *ClOMT2* under different developmental stages of leaves

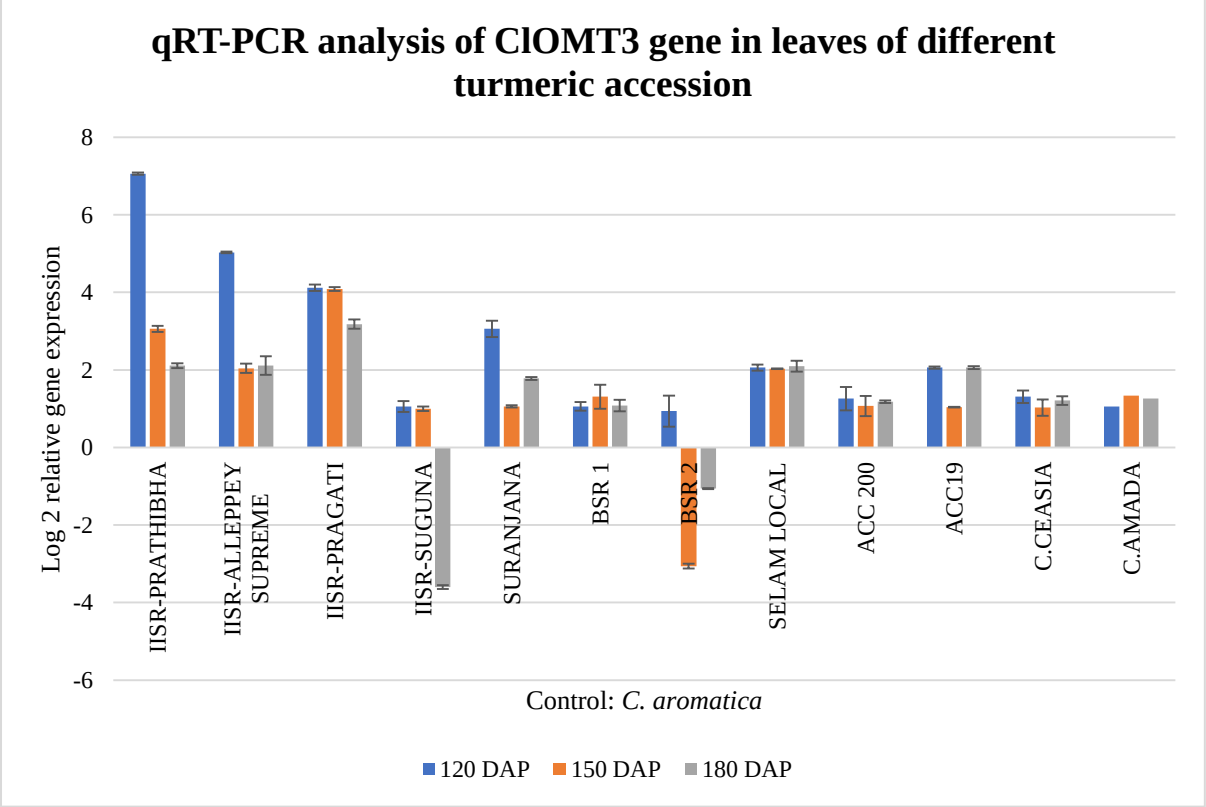


Figure 9: qRT-PCR analysis of *ClOMT3* under different developmental stages of leaves

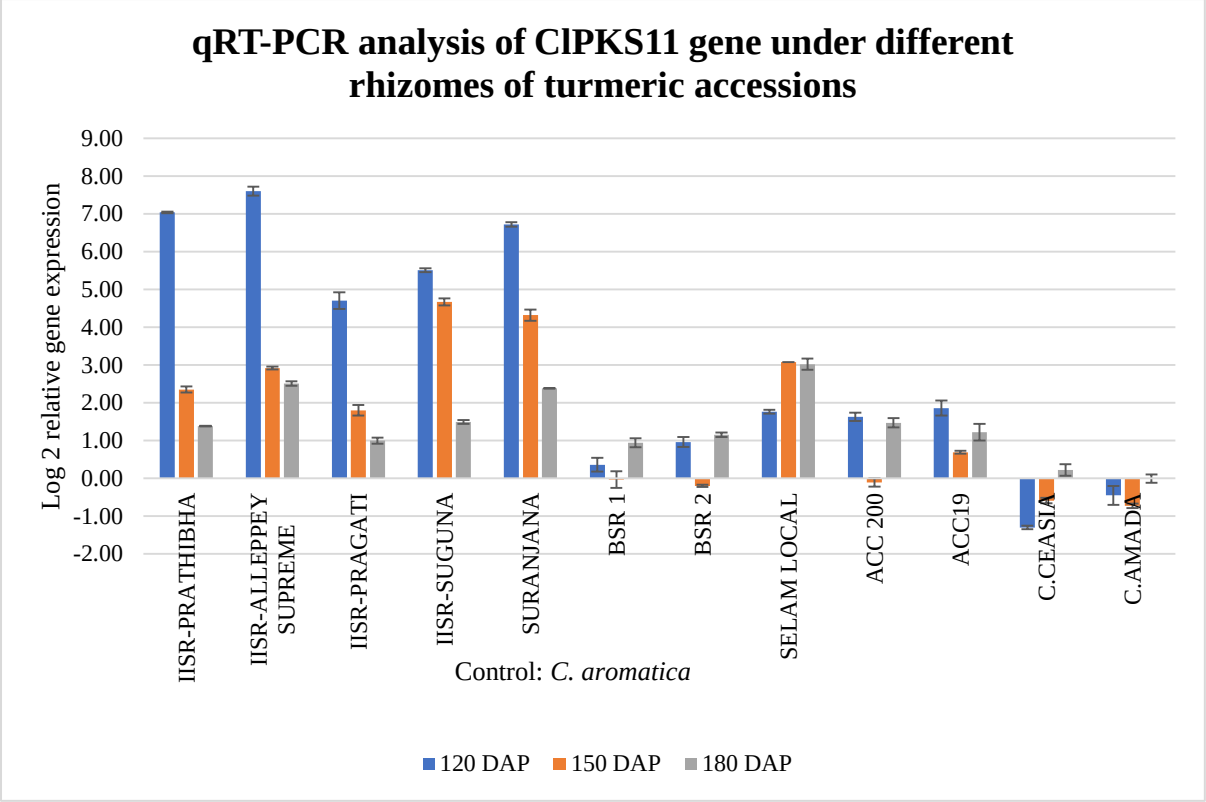


Figure 10: qRT-PCR analysis of *ClPKS11* under different developmental stages of rhizome

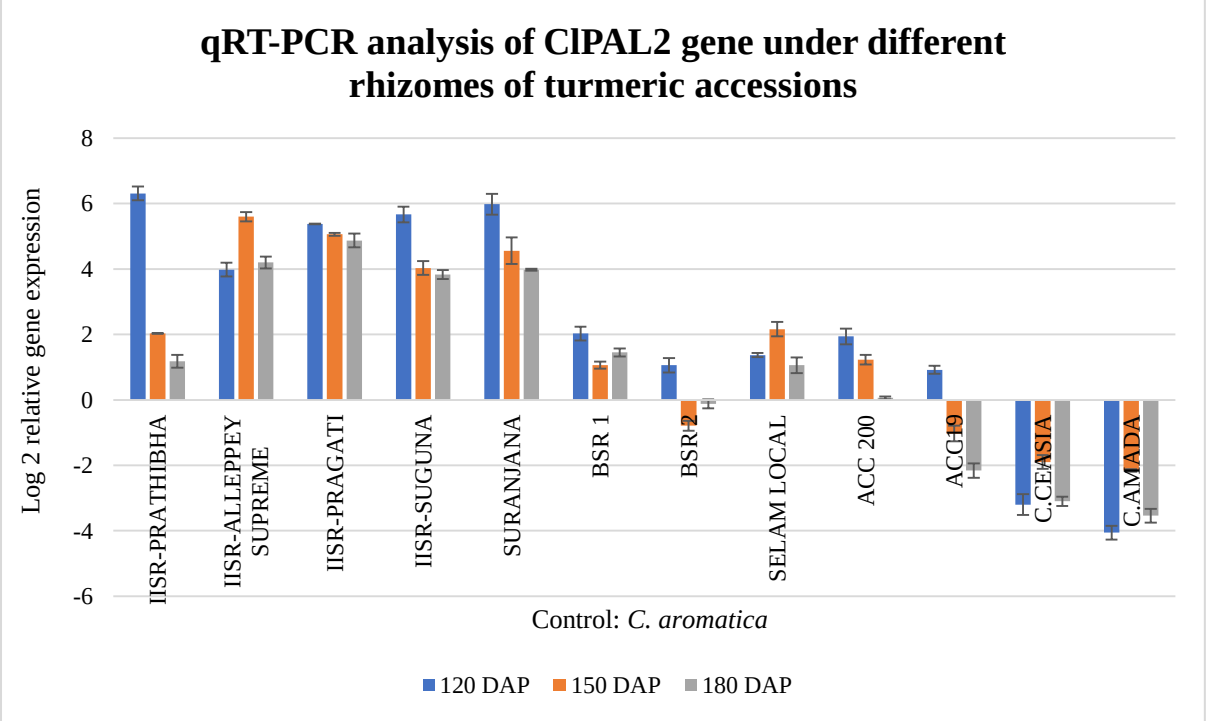


Figure 11: qRT-PCR analysis of *ClPAL2* under different developmental stages of rhizome

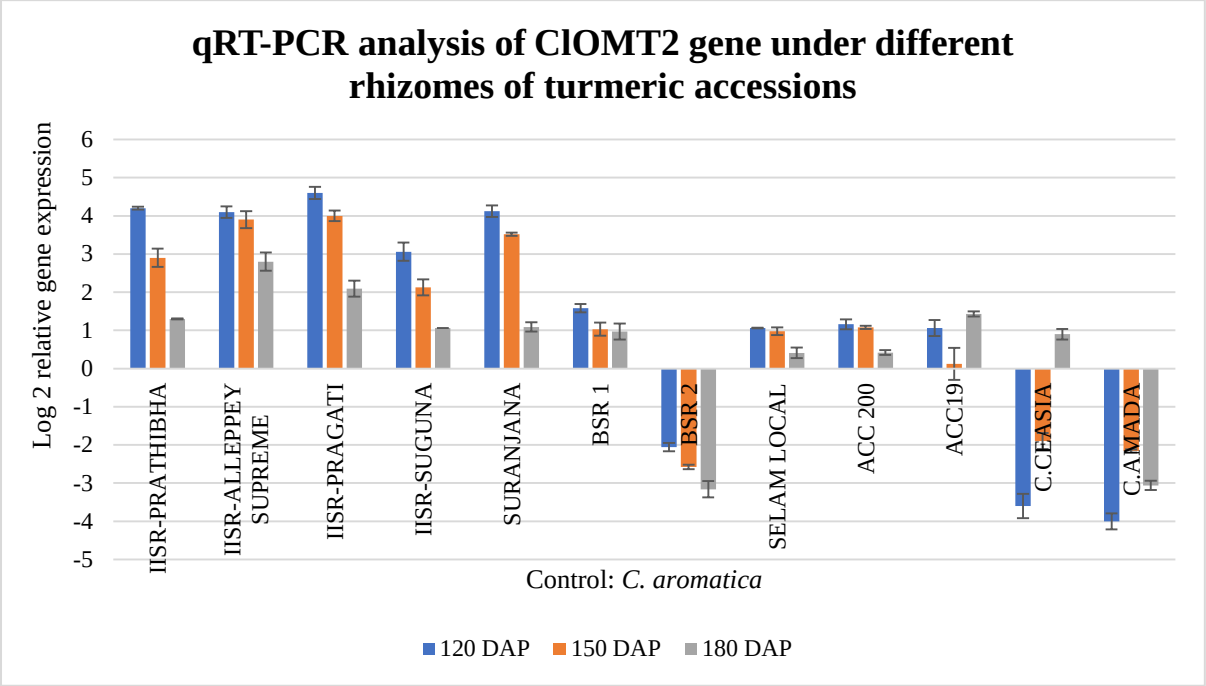


Figure 12: qRT-PCR analysis of *ClOMT2* under different developmental stages of rhizome

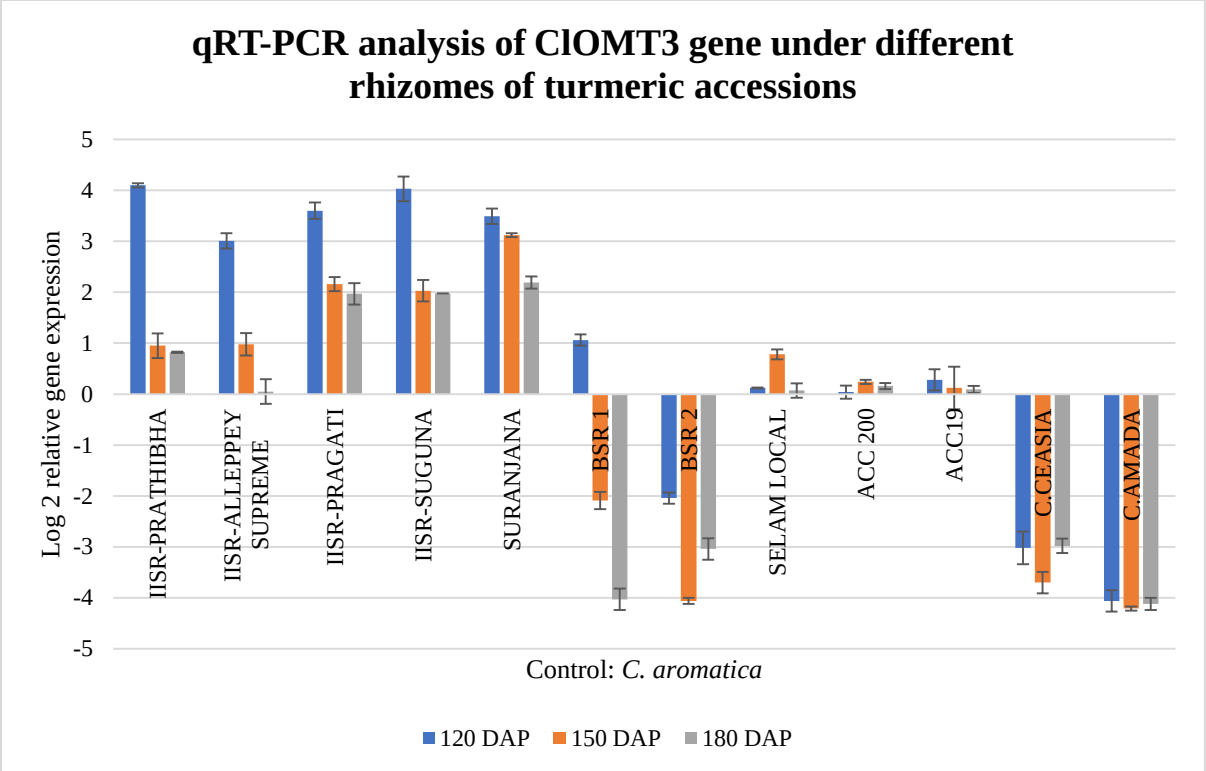


Figure 13: qRT-PCR analysis of *ClOMT3* under different developmental stages of rhizome

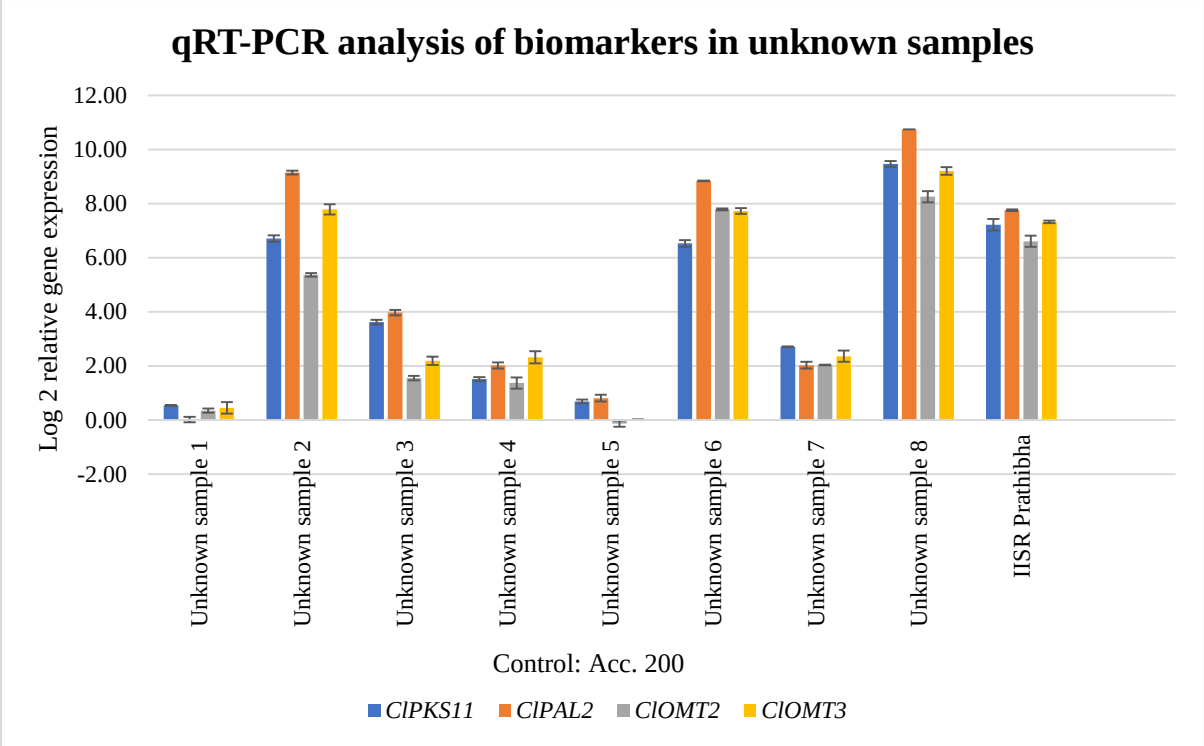


Figure 14: qRT-PCR analysis of biomarkers in unknown samples

4.3.5. Validation of biomarker using blind /unknown samples

In order to select candidate genes as curcumin status markers, we used a non-targeted approach combined with RNA-Seq transcript profiling and qRT PCR based expression analysis. The FPKM of three biomarkers were correlated with their expression levels measured by qRT-PCR, which indicated high degree of consistency between these methods as reported earlier by several workers (Gao et al., 2013; li et al., 2010; 2015; Roberts et al., 2011). Validation of biomarker was done using eight breeding progenies from a population. We used these four biomarkers for screening high and low curcumin lines with IISR Prathibha and Acc.200 as controls. Curcumin status estimated by our method was identical with the curcumin status (high or low) of the blind samples (Figure 14 and Table 3) identified through conventional methods.

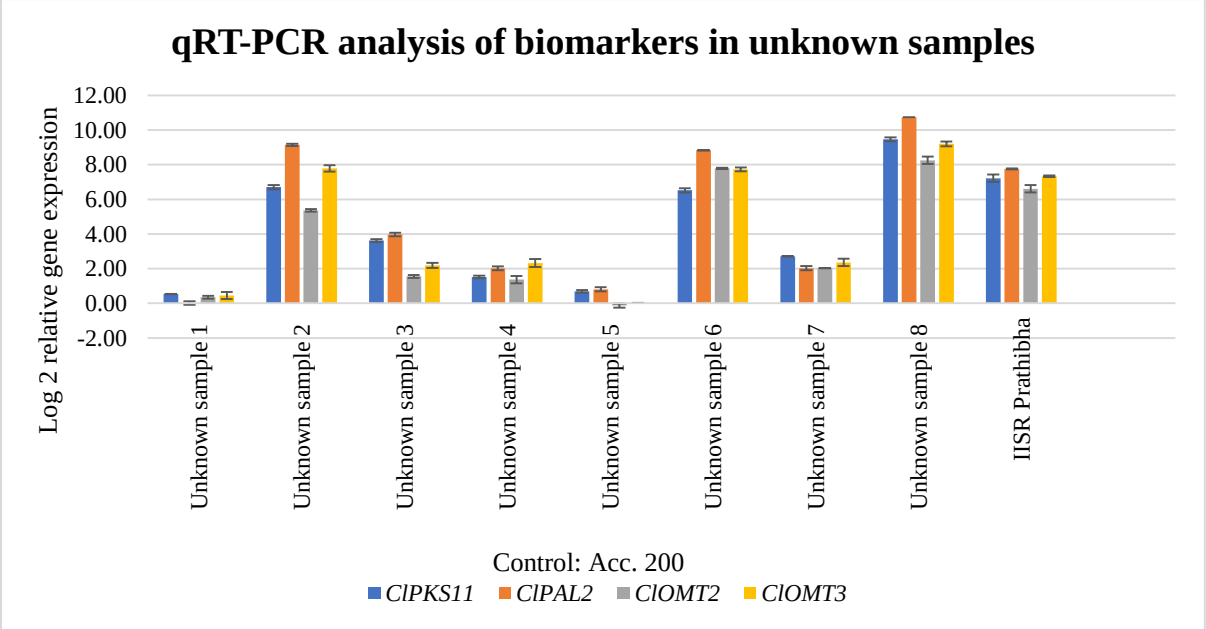


Figure 15: qRT-PCR analysis of biomarkers in unknown samples

Table 3: list of unknown samples with Ct value used for the study

Samples	Ct values				Curcumin status identified by our screening method	Actual curcumin status of the blind samples
	ClPKS11	ClPAL2	ClOMT2	ClOMT3		
Unknown sample 1	28.44	30.06	28.06	29.11	Low curcumin accession	Low curcumin accession
Unknown sample 2	22.37	21.03	23.14	21.87	High curcumin accession	High curcumin accession
Unknown sample 3	25.56	26.31	27.06	27.57	Low curcumin accession	Low curcumin accession
Unknown sample 4	27.56	28.16	27.14	27.34	Low curcumin accession	Low curcumin accession
Unknown sample 5	28.12	29.1	28.37	29.37	Low curcumin accession	Low curcumin accession
Unknown sample 6	22.27	21.06	20.44	21.65	High curcumin accession	High curcumin accession
Unknown sample 7	26.2	27.98	26.3	27.13	Low curcumin accession	Low curcumin accession
Unknown sample 8	19.15	18.97	19.79	19.99	High curcumin accession	High curcumin accession

The method based on the global transcriptome profiling have led to identification of drought markers in potato, Kloosterman et al., 2010, rice (Degenkolbe et al., 2013). However, these studies involve an array of markers and could be probably cumbersome and not economical. In our study we have used only three markers and these markers

are directly involved in the immediate biosynthetic steps following curcumin production and have been validated by several field experiments. These biomarkers showed correlation with curcumin status under all experimental conditions involving different growth and management conditions, tissues, development stages and environments. The use of minimal number of markers makes the strategy less complex and easy to perform especially when large number of samples are analysed.

4.4. CONCLUSION

We developed a qRT-PCR based simple screening strategy for high and low curcumin lines and Ct values of ClPKS11, ClPAL2, ClOMT2 and ClOMT3 used for screening high/low curcumin accessions. It can be concluded that all high curcumin lines may give a Ct value below 22, while low curcumin lines show Ct values above 25. The method is non-destructive, requires very low quantities of plant tissue and can be completed within a day. Methodology is simple and may be used by the pharmaceutical industries, farmers and breeders for screening high/low curcumin lines at early stages of development. The curcumin-based biomarkers may be used singly or together for more accuracy. The technique developed could be employed as a useful tool in different scientific fields for characterising and estimating the curcumin status in plants during development.

- **Prashina Mol P,** Aparna R S, Deepa K and Sheeja T E (2019). "qRT-PCR based screening strategy for curcumin in turmeric (*Curcuma longa* L.)" organized by National Academy of Agricultural Science at Institute of Agricultural Science at Institute of Agricultural Science, Banaras Hindu University, Varanasi, from 13-16 November, 2021

SUMMARY AND FUTURE PERSPECTIVES

Gene expression and co expression studies are very powerful tools in exploring biosynthetic pathways of specialized metabolites. In the present study we have employed only four RNA sequencing samples to shortlist gene candidates regulating biosynthesis of curcumin, which indicates the versatility of expression-based analyses. Over the years our lab has generated a generous amount of information regarding genes and transcripts both under structural as well as regulatory categories of the curcumin biosynthetic pathway. This information along with publicly available data in repositories and the whole genome data available recently has led to uncovering of key genes involved in the pathway and also regulatory molecules like TFs and miRNAs. The results of RNA sequencing showed a significant expression of genes involved in curcumin biosynthesis in high curcumin condition than low curcumin state. Our findings indicated upregulated expression of eleven candidate genes of the pathway, and we have chosen two important genes *ie*., *PAL and OMT* the former being primary upstream regulator and latter one the most important downstream gene immediate to the final metabolite synthesis. Field studies correlating gene expression to curcumin status indicated that high expression levels of these key genes would provide sufficient quantities of precursors for higher curcumin levels. The structural genes identified are novel genes, and the most important achievement of our study is that they are bonafide isoforms involved in the pathway. TFs and miRNAs are excellent study material because manipulating their expression levels can lead to dramatic shifts in metabolite levels. In the study novel bHLH and WD40 TFs and three regulatory miRNAs could be identified that correlated with curcumin in field experiments involving samples with contrasting curcumin. These candidates are advantageous as they can serve as bait genes to uncover new genes, thereby facilitate in exploring new pathways, novel clusters of connected genes *etc*. thus the study has provided valuable insights into major genes and regulatory molecules of the curcumin biosynthetic pathway for engineering high curcumin lines. To further validate the above observations, we examined the iTRAQ-generated proteomic profiles, which helped us in validating a few of the key genes identified through Transcriptomics and co expression studies. These novel genes identified may be utilized as RNA based biomarkers for curcumin status, which may help in easy and fast determination of curcumin status at early developmental stages by qRT-PCR. Integrating omics-based strategies like proteomics and transcriptomics can not only help in validation of genes and proteins, but

can lead to manipulating metabolite levels to suit the needs of farmers and industry. The future research needs to focus on functional validation of candidate genes and TFs identified as putative regulators and unravelling co expression networks, investigations on the TF guided regulation of pathway, promoter mining and regulation, identification of biosynthetic gene clusters and mechanism of regulation and formulating strategies for bio-engineering of curcumin pathway for enhanced biosynthesis and pathway reconstruction.

REFERENCES

Akarchariya, N., Sirilun, S., Julsrigival, J., Chansakaowa, S., 2017. Chemical profiling and antimicrobial activity of essential oil from *Curcuma aeruginosa* Roxb., *Curcuma glans* K. Larsen & J. Mood and *Curcuma cf. xanthorrhiza Roxb*. collected in Thailand. Asian Pac. J. Trop. Biomed. https://doi.org/10.1016/j.apjtb.2017.09.009.

Akbar, A., Kuanar, A., Joshi, R.K., Sandeep, I.S., Mohanty, S., Naik, P.K., Mishra, A., Nayak, S., 2016. Development of prediction model and experimental validation in predicting the curcumin content of turmeric (*Curcuma longa* L.). Front. Plant Sci. https://doi.org/10.3389/fpls.2016.01507.

Akbik, D., Ghadiri, M., Chrzanowski, W., Rohanizadeh, R., 2014. Curcumin as a wound healing agent. Life Sci. https://doi.org/10.1016/j.lfs.2014.08.016.

Akram, M., Ahmed, A., Usmanghani, K., Hannan, A., Mohiuddin, E., Asif, M., 2010. *Curcuma longa* and curcumin: A review article. Rom. J. Biol 55(2), 65-70.

Albert, N.W., Lewis, D.H., Zhang, H., Schwinn, K.E., Jameson, P.E., Davies, K.M., 2011. Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. Plant J. https://doi.org/10.1111/j.1365-313X.2010.04465.

Allwood, E.G., Davies, D.R., Gerrish, C., Ellis, B.E., Bolwell, G.P., 1999. Phosphorylation of phenylalanine ammonia-lyase: Evidence for a novel protein kinase and identification of the phosphorylated residue. FEBS Lett. https://doi.org/10.1016/S0014-5793(99)00998-9.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. https://doi.org/10.1016/S0022-2836(05)80360-2.

Alunni, S., Cipiciani, A., Fioroni, G., Ottavi, L., 2003. Mechanisms of inhibition of phenylalanine ammonia-lyase by phenol inhibitors and phenol/glycine synergistic inhibitors. Arch. Biochem. Biophys. https://doi.org/10.1016/S0003-9861(03)00007-9.

Amato, A., Cavallini, E., Zenoni, S., Finezzo, L., Begheldo, M., Ruperti, B., Tornielli, G.B., 2017. A grapevine TTG2-like WRKY transcription factor is involved in regulating vacuolar transport and flavonoid biosynthesis. Front. Plant Sci. https://doi.org/10.3389/fpls.2016.01979.

Amato, R., Lim, P., Miotto, O., Amaratunga, C., Dek, D., Pearson, R.D., Almagro-Garcia, J., Neal, A.T., Sreng, S., Suon, S., Drury, E., Jyothi, D., Stalker, J., Kwiatkowski, D.P., Fairhurst, R.M., 2017. Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in

Cambodia: a genotype–phenotype association study. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(16)30409-1.

An, J.P., Qu, F.J., Yao, J.F., Wang, X.N., You, C.X., Wang, X.F., Hao, Y.J., 2017. The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. Hortic. Res. https://doi.org/10.1038/hortres.2017.23.

An, X., Jin, G., Luo, X., Chen, C., Li, W., Zhu, G., 2020. Transcriptome analysis and transcription factors responsive to drought stress in *Hibiscus cannabinus*. PeerJ 2020. https://doi.org/10.7717/peerj.8470.

Anand, P., Kunnumakkara, A.B., Newman, R.A., Aggarwal, B.B., 2007. Bioavailability of curcumin: Problems and promises. Mol. Pharm. https://doi.org/ 10. 1021/mp700113r.

Anand, U., Jacobo-Herrera, N., Altemimi, A., Lakhssassi, N., 2019. A comprehensive review on medicinal plants as antimicrobial therapeutics: Potential avenues of biocompatible drug discovery. Metabolites. https://doi.org/ 10.3390/ metabo9110258.

Anandaraj, M., Prasath, D., Kandiannan, K., Zachariah, T.J., Srinivasan, V., Jha, A.K., Singh, B.K., Singh, A.K., Pandey, V.P., Singh, S.P., Shoba, N., Jana, J.C., Kumar, K.R., Maheswari, K.U., 2014. Genotype by environment interaction effects on yield and curcumin in turmeric (*Curcuma longa* L.). Ind. Crops Prod. https://doi.org/10.1016/j.indcrop.2014.01.005.

Andrews, S., 2010. FastQC - A quality control tool for high throughput sequence data. http://www. bioinformatics. babraham. ac.uk/ projects/ fastqc/ Babraham Bioinforma. doi: citeulike-article-id:11583827

Angles, S., Sundar, A., Chinnadurai, M., 2011. Impact of Globalization on Production and Export of Turmeric in India – An Economic Analysis 24, 301–308.

Annadurai, R.S., Neethiraj, R., Jayakumar, V., Damodaran, A.C., Rao, S.N., Katta, M.A.V.S.K., Gopinathan, S., Sarma, S.P., Senthilkumar, V., Niranjan, V., Gopinath, A., 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. https://doi.org/ 10.1371/ journal. pone. 0056217.

Annis, J., Coons, J., Zaya, D.N., Molano-Flores, B., 2019. Environmental influences on foliar anthocyanin production in *Pinguicula planifolia* (Lentibulariaceae). J. Torrey Bot. Soc. 146, 269–277. https://doi.org/10.3159/TORREY-D-18-00043.1.

Aoki, K., Ogata, Y., Shibata, D., 2007. Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol. https://doi.org/10.1093/pcp/pcm013.

Appert, C., Logemann, E., Hahlbrock, K., Schmid, J., Amrhein, N., 1994. Structural and Catalytic Properties of the Four Phenylalanine Ammonia Lyase Isoenzymes from Parsley (*Petroselinum Crispum* Nym.). Eur. J. Biochem. https://doi.org/10.1111/j.1432-1033.1994.00491.

Araújo, C.A.C., Leon, L.L., 2001. Biological activities of *Curcuma longa* L. Mem. Inst. Oswaldo Cruz. https://doi.org/10.1590/S0074-02762001000500026.

Arct, J., Ratz-Lyko, A., Mieloch, M., Witulska, M., 2014. Evaluation of skin colouring properties of *Curcuma longa* extract. Indian J. Pharm. Sci 76(4), 374.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics. https://doi.org/10.1093/bioinformatics/bti770.

Arrell, D.K., Neverova, I., Van Eyk, J.E., 2001. Cardiovascular proteomics evolution and potential. Circ. Res. https://doi.org/10.1161/hh0801.090193.

Arulkumar, A., Ramanchandran, K., Paramasivam, S., Palanivel, R., Miranda, J.M., 2017. Effects of turmeric (*Curcuma longa*) on shelf-life extension and biogenic amine control of cuttlefish (*Sepia brevimana*) during chilled storage. CyTA - J. Food. https://doi.org/10.1080/19476337.2017.1296495.

Aslam, B., Basit, M., Nisar, M.A., Khurshid, M., Rasool, M.H., 2017. Proteomics: Technologies and their applications. J. Chromatogr. Sci. https:// doi.org/10.1093/chromsci/bmw167.

Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., Orhan, I.E., Banach, M., Rollinger, J.M., Barreca, D., Weckwerth, W., Bauer, R., Bayer, E.A., Majeed, M., Bishayee, A., Bochkov, V., Bonn, G.K., Braidy, N., Bucar, F., Cifuentes, A., D'Onofrio, G., Bodkin, M., Diederich, M., Dinkova-Kostova, A.T., Efferth, T., El Bairi, K., Arkells, N., Fan, T.P., Fiebich, B.L., Freissmuth, M., Georgiev, M.I., Gibbons, S., Godfrey, K.M., Gruber, C.W., Heer, J., Huber, L.A., Ibanez, E., Kijjoa, A., Kiss, A.K., Lu, A., Macias, F.A., Miller, M.J.S., Mocan, A., Müller, R., Nicoletti, F., Perry, G., Pittalà, V., Rastrelli, L., Ristow, M., Russo, G.L., Silva, A.S., Schuster, D., Sheridan, H., Skalicka-Wo?niak, K., Skaltsounis, L., Sobarzo-Sánchez, E., Bredt, D.S., Stuppner, H., Sureda, A., Tzvetkov, N.T., Vacca, R.A., Aggarwal, B.B., Battino, M., Giampieri, F., Wink, M., Wolfender, J.L., Xiao, J., Yeung, A.W.K., Lizard, G., Popp, M.A., Heinrich, M., Berindan-Neagoe, I., Stadler, M., Daglia, M., Verpoorte, R., Supuran, C.T., 2021. Natural products in drug discovery: advances and opportunities. Nat. Rev. Drug Discov. https://doi.org/10.1038/s41573-020-00114-z.

Atchley, W.R., Terhalle, W., Dress, A., 1999. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J. Mol. Evol. https://doi.org/10.1007/ PL00006494.

Azuma, A., 2018. Genetic and environmental impacts on the biosynthesis of anthocyanins in grapes. Hortic. J. 87, 1–17. https://doi.org/10.2503/HORTJ.OKD-IR02.

Babaei, F., Nassiri-Asl, M., Hosseinzadeh, H., 2020. Curcumin (a constituent of turmeric): New treatment option against COVID-19. Food Sci. Nutr. https://doi.org/10.1002/fsn3.1858.

Babu, N., 2015. Traditional Cultivation Practices of Turmeric in Tribal Belt of Odisha. J. Eng. Comput. Appl. Sci. 4(2), 52-57.

Bachem, C.W.B., Van Der Hoeven, R.S., De Bruijn, S.M., Vreugdenhil, D., Zabeau, M., Visser, R.G.F., 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J. https://doi.org/10.1046/j.1365-313X.1996.9050745.

Bagal, U.R., Leebens-Mack, J.H., Lorenz, W.W., Dean, J.F.D., 2012. The phenylalanine ammonia lyase (*PAL*) gene family shows a gymnosperm-specific lineage. BMC Genomics. https://doi.org/10.1186/1471-2164-13-s3-s1.

Bao Ting Zhu, Ezell, E.L., Liehr, J.G., 1994. Catechol-O-methyltransferase-catalyzed rapid O-methylation of mutagenic flavonoids. Metabolic inactivation as a possible reason for their lack of carcinogenicity *in vivo*. J. Biol. Chem. https://doi.org/10.1016/s0021-9258(17) 42348-9.

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L., Schnable, P.S., 2007. SNP discovery *via* 454 transcriptome sequencing. Plant J. https://doi.org/10.1111/j.1365-313X.2007.03193.

Barros, J., Dixon, R.A., 2020. Plant Phenylalanine/Tyrosine Ammonia-lyases. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2019.09.011.

Basnet, P., Skalko-Basnet, N., 2011. Curcumin: An anti-inflammatory molecule from a curry spice on the path to cancer treatment. Molecules. https://doi.org/10.3390/molecules16064567.

Bate, N.J., Orr, J., Ni, W., Meromi, A., Nadler-Hassar, T., Doerner, P.W., Dixon, R.A., Lamb, C.J., Elkind, Y., 1994. Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate- determining step in natural product synthesis. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.91.16.7608.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L.L., 2002. The pfam protein families database. Nucleic Acids Res. https://doi.org/10.1093/nar/30.1.276.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R.,

Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Gane, P., Castro, L.G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W., Luo, J., Macdougall, A., Mutowo, P., Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Cowley, A., Figueira, L., Li, W., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Suzek, B.E., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Yerramalla, M.S., Zhang, J., 2015. UniProt: A hub for protein information. Nucleic Acids Res. https://doi.org/10.1093/nar/gku989.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., 2021. UniProt: The

universal protein knowledgebase in 2021. Nucleic Acids Res. https://doi.org/10.1093/nar/gkaa1100.

Baxter, H.L., Stewart, C.N., 2013. Effects of altered lignin biosynthesis on phenylpropanoid metabolism and plant stress. Biofuels. https://doi.org/10.4155/bfs.13.56 https://doi.org/10.4155/bfs.13.56

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins Struct. Funct. Genet. https://doi.org/10.1002/prot.21715.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. https://doi.org/10.1093/nar/28.1.235.

Bhat, W.W., Dhar, N., Razdan, S., Rana, S., Mehra, R., Nargotra, A., Dhar, R.S., Ashraf, N., Vishwakarma, R., 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. https://doi.org/10.1371/journal.pone.0073804.

Bhowmik, D., Kumar, K.P., Chandira, M., Jayakar, B., Jit, 2009. Turmeric: A Herbal and Traditional Medicine. Arch. Appl. Sci. Res 1 (2) 86-108.

Bhuiya, M.W., Liu, C.J., 2010. Engineering monolignol 4-O-methyltransferases to modulate lignin biosynthesis. J. Biol. Chem. https://doi.org/10.1074/jbc.M109.036673

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. J. Exp. Bot. https://doi.org/10.1093/jxb/ern290

Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., 2017. The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw1132.

Biswas, S., Hazra, S., Chattopadhyay, S., 2016. Identification of conserved miRNAs and their putative target genes in *Podophyllum hexandrum* (Himalayan Mayapple). Plant Gene. https://doi.org/10.1016/j.plgene.2016.04.002.

Bolwell, G.P., 1992. A role for phosphorylation in the down-regulation of phenylalanine ammonia-lyase in suspension-cultured cells of french bean. Phytochemistry. https://doi.org/10.1016/0031-9422(92)80418-E.

Bolwell, G.P., Bell, J.N., Cramer, C.L., Schuch, W., Lamb, C.J., Dixon, R.A., 1985. L Phenylalanine ammonia lyase from *Phaseolus vulgaris* Characterisation and differential induction of multiple forms from elicitor treated cell suspension cultures. Eur. J. Biochem. https://doi.org/10.1111/j.1432-1033.1985.tb08941.

Bolwell, G.P., Cramer, C.L., Lamb, C.J., Schuch, W., Dixon, R.A., 1986. L-Phenylalanine ammonia-lyase from *Phaseolus vulgaris*: Modulation of the levels of active enzyme by trans-cinnamic acid. Planta. https://doi.org/10.1007/BF01369780.

Bourgaud, F., Gravot, A., Milesi, S., Gontier, E., 2001. Production of plant secondary metabolites: A historical perspective. Plant Sci. https://doi.org/10.1016/S0168-9452(01)00490-3.

Breitling, R., Ceniceros, A., Jankevics, A., Takano, E., 2013. Metabolomics for Secondary Metabolite Research. Metabolites. https://doi.org/10.3390/metabo3041076.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., Corcoran, K., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. 18, 630–634. https://doi.org/10.1038/76469.

Bruscella, P., Bottini, S., Baudesson, C., Pawlotsky, J.M., Feray, C., Trabucchi, M., 2017. Viruses and miRNAs: More friends than foes. Front. Microbiol. https://doi.org/10.3389/fmicb.2017.00824.

Bureau, T., Lam, K.C., Ibrahim, R.K., Behdad, B., Dayanandan, S., 2007. Structure, function, and evolution of plant O-methyltransferases. Genome. https://doi.org/10.1139/G07-077.

Busam, G., Junghanns, K.T., Kneusel, R.E., Kassemeyer, H.H., Matern, U., 1997. Characterization and expression of caffeoyl-coenzyme a 3-O-methyltransferase proposed for the induced resistance response of *Vitis vinifera* L. Plant Physiol. https://doi.org/10.1104/pp.115.3.1039.

Calabrese, J.C., Jordan, D.B., Boodhoo, A., Sariaslani, S., Vannelli, T., 2004. Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis. Biochemistry. https://doi.org/10.1021/bi049053+

Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A.A., Paiva, J.A.P., Grima-Pettenati, J., 2015. Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. New Phytol. https://doi.org/10.1111/nph.13313.

Cas, M.D., Ghidoni, R., 2019. Dietary curcumin: Correlation between bioavailability and health potential. Nutrients. https://doi.org/10.3390/nu11092147.

Casado-Vela, J., Martínez-Esteso, M.J., Rodriguez, E., Borrás, E., Elortza, F., Bru-Martínez, R., 2010. iTRAQ-based quantitative analysis of protein mixtures with large fold change and dynamic range. Proteomics 10, 343–347. https:// doi.org/ 10.1002/PMIC.200900509.

Catalanotto, C., Cogoni, C., Zardo, G., 2016. MicroRNA in control of gene expression: An overview of nuclear functions. Int. J. Mol. Sci. https://doi.org/ 10.3390/ijms17101712.

Celis, J.E., Celis, P., Østergaard, M., Basse, B., Lauridsen, J.B., Ratz, G., Rasmussen, H.H., Ørntoft, T.F., Hein, B., Wolf, H., Celis, A., 1999. Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: A novel strategy for identifying metaplastic lesions. Cancer Res 59(12):3003-3009.

Chabrillange, N., Talamond, P., Moreau, C., Le Gal, L., Bourgeois, M., Hamon, S., Noirot, M., De, A., Campa, C., 2006. Isolation and First Characterization of Two O-Methyltransferase Genes Involved in Phenylpropanoid Pathway in *Coffea canephora*. 10.1007/s00425-012-1613-2.

Chagné, D., Lin-Wang, K., Espley, R. V., Volz, R.K., How, N.M., Rouse, S., Brendolise, C., Carlisle, C.M., Kumar, S., de Silva, N., Micheletti, D., McGhie, T., Crowhurst, R.N., Storey, R.D., Velasco, R., Hellens, R.P., Gardiner, S.E., Allan, A.C., 2013. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotyp. Plant Physiol. https://doi.org/10.1104/pp.112.206771.

Chainani-Wu, N., 2003. Safety and anti-inflammatory activity of curcumin: A component of tumeric (*Curcuma longa*). J. Altern. Complement. Med. 10.1089/107555303321223035

Chakraborty, A., Mahajan, S., Jaiswal, S.K., Sharma, V.K., 2021. Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. Commun. Biol. https://doi.org/10.1038/s42003-021-02720.

Chalker-Scott, L., 1999. Environmental Significance of Anthocyanins in Plant Stress Responses. Photochem. Photobiol. 70, 1–9. https://doi.org/10.1111/J.1751-1097.1999.TB01944.

Chandra, H., Bishnoi, P., Yadav, A., Patni, B., Mishra, A.P., Nautiyal, A.R., 2017. Antimicrobial resistance and the alternative resources with special emphasis on plant-based antimicrobials - A review. Plants. https://doi.org/ 10.3390/ plants6020016.

Chang, A., Lim, M.H., Lee, S.W., Robb, E.J., Nazar, R.N., 2008. Tomato phenylalanine ammonia-lyase gene family, highly redundant but strongly underutilized. J. Biol. Chem. https://doi.org/10.1074/jbc.M804428200.

Chang, L., Hagel, J.M., Facchini, P.J., 2015. Isolation and characterization of O-methyltransferases involved in the biosynthesis of glaucine in *Glaucium flavum*. Plant Physiol. 169, 1127–1140. https://doi.org/10.1104/pp.15.01240.

Chang, T.S., Liu, C.W., Lin, Y.L., Li, C.Y., Wang, A.Z., Chien, M.W., Wang, C.S., Lai, C.C., 2017. Mapping and comparative proteomic analysis of the starch

biosynthetic pathway in rice by 2D PAGE/MS. Plant Mol. Biol. https://doi.org/10.1007/s11103-017-0652-2.

Chattopadhyay, I., Biswas, K., Bandyopadhyay, U., Banerjee, R.K., 2004b. Turmeric and curcumin: Biological actions and medicinal applications. *Curr. Sci* 10, 44-55.

Chempakam, B., and Parthasarathy, V. A. (2008). Turmeric. Chemistry of Spices 6, 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, e0159238. https://doi.org/10.1371/JOURNAL.PONE.0159238.

Chen, L., Lu, D., Sun, K., Xu, Y., Hu, P., Li, X., Xu, F., 2019. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene. https://doi.org/ 10.1016/ j. gene. 2019.01.001

Chen, L., Wu, F., Zhang, J., 2021. Nac and myb families and lignin biosynthesis-related members identification and expression analysis in *Melilotus albus*. Plants. https://doi.org/10.3390/plants10020303.

Chen, L., Yuan, L., Qian, K., Qian, G., Zhu, Y., Wu, C.L., Dan, H.C., Xiao, Y., Wang, X., 2018. Identification of biomarkers associated with pathological stage and prognosis of clear cell renal cell carcinoma by co-expression network analysis. Front. Physiol. https://doi.org/10.3389/fphys.2018.00399.

Chen, Y.Z., Pang, Q.Y., He, Y., Zhu, N., Branstrom, I., Yan, X.F., Chen, S., 2012. Proteomics and metabolomics of arabidopsis responses to perturbation of glucosinolate biosynthesis. Mol. Plant. https://doi.org/10.1093/mp/sss034.

Chen, Z., Zheng, Z., Huang, J., Lai, Z., Fan, B., 2009. Biosynthesis of salicylic acid in plants. Plant Signal. Behav. https://doi.org/10.4161/psb.4.6.8392.

Cheng, S.H., Sheen, J., Gerrish, C., Bolwell, G.P., 2001. Molecular identification of phenylalanine ammonia-lyase as a substrate of a specific constitutively active Arabidopsis CDPK expressed in maize protoplasts. FEBS Lett. https:// doi.org/ 10.1016/S0014-5793(01)02732-6.

Choe, L., D'Ascenzo, M., Relkin, N.R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., Lee, K.H., 2007. 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics. https://doi.org/10.1002/pmic.200700316.

Chunthaburee, S., Sakuanrungsirikul, S., Wongwarat, T., Sanitchon, J., Pattanagul, W., Theerakulpisut, P., 2016. Changes in anthocyanin content and expression of anthocyanin synthesis genes in seedlings of black glutinous rice in response to salt stress. Asian J. Plant Sci. https://doi.org/10.3923/ajps.2016.56.65.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., Bruce, S.J., Lee, C.C., Ranade, S.S., Peckham, H.E., Manning, J.M., McKernan, K.J., Grimmond, S.M., 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods. https://doi.org/10.1038/nmeth.1223.

Cochrane, F.C., Davin, L.B., Lewis, N.G., 2004. The Arabidopsis phenylalanine ammonia lyase gene family: Kinetic characterization of the four PAL isoforms. Phytochemistry. https://doi.org/10.1016/j.phytochem.2004.05.006.

Conesa, A., Götz, S., 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics. https://doi.org/10.1155/2008/619832.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. https://doi.org/10.1093/bioinformatics/bti610.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. https://doi.org/ 10.1093 /BIOINFORMATICS/BTI610.

Cramer, C.L., Edwards, K., Dron, M., Liang, X., Dildine, S.L., Bolwell, G.P., Dixon, R.A., Lamb, C.J., Schuch, W., 1989. Phenylalanine ammonia-lyase gene organization and structure. Plant Mol. Biol. https://doi.org/10.1007/BF00017577.

Cui, D., Zhao, S., Xu, H., Allan, A.C., Zhang, X., Fan, L., Chen, L., Su, J., Shu, Q., Li, K., 2021. The interaction of MYB, bHLH and WD40 transcription factors in red pear (*Pyrus pyrifolia*) peel. Plant Mol. Biol. https://doi.org/10.1007/s11103-021-01160

Dassanayake, M., Haas, J.S., Bohnert, H.J., Cheeseman, J.M., 2009. Shedding light on an extremophile lifestyle through transcriptomics. New Phytol. https://doi.org/ 10.1111/j.1469-8137.2009.02913.

Daube F. W, 1870. Ueber Curcumin, den Farbstoff der *Curcuma*wurzel. Berichte Der Dtsch. Chem. Gesellschaft 3, 609–613.

De Paolis, A., Caretto, S., Quarta, A., Di Sansebastiano, G. Pietro, Sbrocca, I., Frugis, G., Mita, G., 2020. Genome-wide identification of wrky genes in *Artemisia annua*: Characterization of a putative ortholog of ATWRKY40. Plants. https://doi.org/ 10.3390/plants9121669.

Deepa, K. 2018. Transcriptome and proteome analysis of the curcuminoid biosynthetic pathway in turmeric (*Curcuma longa* L.). PhD thesis. Calicut University.

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.). Physiol. Mol. Biol. Plants. https://doi.org/10.1007/s12298-013-0218-y.

Deepa, K., Sheeja, T.E., Rosana, O.B., Srinivasan, R., Krishnamoorthy K.S., Sasikumar, B., 2017. Highly conserved sequence of ClPKS11 encodes a novel polyketide synthase involved in curcumin biosynthesis in turmeric (*Curcuma longa* L.). Ind. Crops Prod. https://doi.org/10.1016/j.indcrop.2016.12.003.

Degenkolbe, T., Do, P.T., Kopka, J., Zuther, E., Hincha, D.K., Köhl, K.I., 2013. Identification of Drought Tolerance Markers in a Diverse Population of Rice Cultivars by Expression and Metabolite Profiling. PLoS One. https://doi.org/10.1371/journal.pone.0063637.

Deng, Y., Lu, S., 2017. Biosynthesis and Regulation of Phenylpropanoids in Plants. CRC. Crit. Rev. Plant Sci. https://doi.org/10.1080/07352689.2017.1402852.

Diray-Arce, J., Clement, M., Gul, B., Khan, M.A., Nielsen, B.L., 2015. Transcriptome assembly, profiling and differential gene expression analysis of the halophyte *Suaeda fruticosa* provides insights into salt tolerance. BMC Genomics. https://doi.org/10.1186/s12864-015-1553-x.

Dixit, D., Srivastava, N.K., Sharma, S., 2002. Boron deficiency induced changes in translocation of $14CO_2$-photosynthate into primary metabolites in relation to essential oil and curcumin accumulation in turmeric (*Curcuma longa* L.). Photosynthetica. https://doi.org/10.1023/A:1020118913452.

Dixon, R.A., Paiva, N.L., 1995. Stress-induced phenylpropanoid metabolism. Plant Cell. https://doi.org/10.1105/tpc.7.7.1085.

Djami-Tchatchou, A.T., Sanan-Mishra, N., Ntushelo, K., Dubery, I.A., 2017. Functional roles of microRNAs in agronomically important plants-potential as targets for crop improvement and protection. Front. Plant Sci. https://doi.org/ 10.3389/ fpls.2017.00378.

Dong, C.J., Shang, Q.M., 2013. Genome-wide characterization of phenylalanine ammonia-lyase gene family in watermelon (*Citrullus lanatus*). Planta. https://doi.org/10.1007/s00425-013-1869-1.

Dong, N.Q., Lin, H.X., 2021. Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. J. Integr. Plant Biol. https://doi.org/10.1111/jipb.13054.

Dong, T., Zhu, M., Yu, J., Han, R., Tang, C., Xu, T., Liu, J., Li, Z., 2019. RNA-Seq and iTRAQ reveal multiple pathways involved in storage root formation and development in sweet potato (*Ipomoea batatas* L.). BMC Plant Biol. 19, 1–16. https://doi.org/10.1186/s12870-019-1731-0.

Dong, Z.C., Chen, Y., 2013. Transcriptomics: Advances and approaches. Sci. China Life Sci. https://doi.org/10.1007/s11427-013-4557-2.

Dosoky, N.S., Setzer, W.N., 2018. Chemical composition and biological activities of essential oils of *curcuma* species. Nutrients. https://doi.org/10.3390/nu10091196.

Dowle, A.A., Wilson, J., Thomas, J.R., 2016. Comparing the Diagnostic Classification Accuracy of iTRAQ, Peak-Area, Spectral-Counting, and emPAI Methods for Relative Quantification in Expression Proteomics. J. Proteome Res. 15, 3550–3562.https://doi.org/10.1021/ACS.JPROTEOME.6B00308/SUPPL_FILE/PR6B00308_SI_007.XLSX.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., Lepiniec, L., 2010. MYB transcription factors in Arabidopsis. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2010.06.005.

Durbin, M.L., Learn, G.H., Huttley, G.A., Clegg, M.T., 1995. Evolution of the chalcone synthase gene family in the genus Ipomoea. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.92.8.3338.

Durbin, M.L., McCaig, B., Clegg, M.T., 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. https://doi.org/10.1023/A:1006375904820.

Edwards, D., Batley, J., Snowdon, R.J., 2013. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. https://doi.org/10.1007/s00122-012-1964-x.

Elliott, M.H., Smith, D.S., Parker, C.E., Borchers, C., 2009. Current trends in quantitative proteomics. J. Mass Spectrom. 44, 1637–1660. https://doi.org/10.1002/jms.1692

Estabrook, E.M., Sengupta-Gopalan, C., 1991. Differential expression of phenylalanine ammonia-lyase and chalcone synthase during soybean nodule development. Plant Cell. https://doi.org/10.1105/tpc.3.3.299.

Evans, C., Noirel, J., Ow, S.Y., Salim, M., Pereira-Medrano, A.G., Couto, N., Pandhal, J., Smith, D., Pham, T.K., Karunakaran, E., Zou, X., Biggs, C.A., Wright, P.C., 2012. An insight into iTRAQ: Where do we stand now? Anal. Bioanal. Chem. https://doi.org/10.1007/s00216-012-5918-6.

Ewon, K., Bhagya, A.S., 2019. A review on golden species of Zingiberaceae family around the world: Genus *Curcuma*. African J. Agric. Res. https://doi.org/10.5897/ajar2018.13755.

Fan, J., Chen, C., Yu, Q., Brlansky, R.H., Li, Z.G., Gmitter, F.G., 2011. Comparative iTRAQ proteome and transcriptome analyses of sweet orange infected by "*Candidatus Liberibacter asiaticus*." Physiol. Plant. 143, 235–245. https://doi.org/10.1111/j.1399-3054.2011.01502-x.

Farrow, S.C., Kamileen, M.O., Meades, J., Ameyaw, B., Xiao, Y., O'Connor, S.E., 2018. Cytochrome P450 and O-methyltransferase catalyze the final steps in the

biosynthesis of the anti-addictive alkaloid ibogaine from *Tabernanthe iboga*. J. Biol. Chem. https://doi.org/10.1074/jbc.RA118.004060.

Favia, A. D., Nobeli, I., Glaser, F., and Thornton, J. M. 2008. Molecular docking for substrate identification: the short-chain dehydrogenases/reductases. *Journal of mol. bio, 375* (3), 855-874.

Ferrer, J.L., Zubieta, C., Dixon, R.A., Noel, J.P., 2005. Crystal structures of alfalfa caffeoyl coenzyme A 3-O-methyltransferase. Plant Physiol. https://doi.org/ 10.1104/pp.104.048751.

Fofana, B., Somalraju, A., Fillmore, S., Zaidi, M., Main, D., Ghose, K., 2020. Comparative transcriptome expression analysis in susceptible and resistant potato (*Solanum tuberosum*) cultivars to common scab (*Streptomyces scabies*) revealed immune priming responses in the incompatible interaction. PLoS One 15, 1–27. https://doi.org/10.1371/journal.pone.0235018.

Foong, L.C., Chai, J.Y., Ho, A.S.H., Yeo, B.P.H., Lim, Y.M., Tam, S.M., 2020. Comparative transcriptome analysis to identify candidate genes involved in 2-methoxy-1,4-naphthoquinone (MNQ) biosynthesis in *Impatiens balsamina* L. Sci. Rep. https://doi.org/10.1038/s41598-020-72997-2.

Fraser, C.M., Chapple, C., 2011. The Phenylpropanoid Pathway in Arabidopsis. Arab. B. https://doi.org/10.1199/tab.0152.

Fritz, R.R., Hodgins, D.S., Abell, C.W., 1976. Phenylalanine ammonia-lyase. Induction and purification from yeast and clearance in mammals. J. Biol. Chem. https://doi.org/10.1016/s0021-9258(17)33251-9.

Frohman, M.A., 1994. On beyond classic RACE (Rapid Amplification of cDNA Ends). PCR Methods Appl. https://doi.org/10.1101/gr.4.1.s40.

Ganapathy, G., Keerthi, D., Nair, R.A., Pillai, P., 2016. Correlation of Phenylalanine ammonia lyase (PAL) and Tyrosine ammonia lyase (TAL) activities to phenolics and curcuminoid content in ginger and its wild congener, *Zingiber zerumbet* following *Pythium myriotylum* infection. Eur. J. Plant Pathol. https://doi.org/ 10.1007/ s10658-016-0865-2.

Gao, J., Li, W.B., Liu, H.F., Chen, F.B., 2020. Identification of differential expression genes related to anthocyanin biosynthesis in carmine radish (*Raphanus sativus* L.) fleshy roots using comparative RNA-Seq method. PLoS One. https://doi.org/10.1371/journal.pone.0231729.

Gao, S., Wang, F., Niran, J., Li, N., Yin, Y., Yu, C., Jiao, C., Yao, M., 2021. Transcriptome analysis reveals defenserelated genes and pathways against *Xanthomonas campestris* pv. vesicatoria in pepper (*Capsicum annuum* L.). PLoS One. https://doi.org/10.1371/journal.pone.0240279.

Gaudet, P., Logie, C., Lovering, R.C., Kuiper, M., Lægreid, A., Thomas, P.D., 2021. Gene Ontology representation for transcription factor functions. Biochim. Biophys. Acta - Gene Regul. Mech. 1864, 194752. https://doi.org/10.1016/ J. BBAGRM. 2021.194752.

Gebhardt, C., 2013. Bridging the gap between genome analysis and precision breeding in potato. Trends Genet. https://doi.org/10.1016/j.tig.2012.11.006.

Geourjon, C., Deléage, G., 1995. Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. https://doi.org/10.1093/bioinformatics/11.6.681.

Gera, M., Sharma, N., Ghosh, M., Huynh, D.L., Lee, S.J., Min, T., Kwon, T., Jeong, D.K., 2017. Nanoformulations of curcumin: An emerging paradigm for improved remedial application. Oncotarget. https://doi.org/10.18632/oncotarget.19164.

Giordano, D., Provenzano, S., Ferrandino, A., Vitali, M., Pagliarani, C., Roman, F., Cardinale, F., Castellarin, S.D., Schubert, A., 2016. Characterization of a multifunctional caffeoyl-CoA O-methyltransferase activated in grape berries upon drought stress. Plant Physiol. Biochem. 101, 23–32. https://doi.org/ 10.1016/J.PLAPHY.2016.01.015.

Gonçalves, G.M.S., Da Silva, G.H., Barros, P.P., Srebernich, S.M., Shiraishi, C.T.C., De Camargos, V.R., Lasca, T.B., 2014. Use of *Curcuma longa* in cosmetics: Extraction of curcuminoid pigments, development of formulations, and *in vitro* skin permeation studies. Brazilian J. Pharm. Sci. 50, 885–894. https://doi.org/ 10.1590/ S1984-82502014000400024.

Gong, W., Xu, F., Sun, J., Peng, Z., He, S., Pan, Z., Du, X., 2017. ITRAQ-based comparative proteomic analysis of seedling leaves of two upland cotton genotypes differing in salt tolerance. Front. Plant Sci. 8, 2113. https://doi.org/10.3389/ FPLS. 2017.02113/BIBTEX.

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. *The Plant Journal*, **53**(5): 814-827.

Gopinath, H., Karthikeyan, K., 2018. Turmeric: A condiment, cosmetic and cure. Indian J. Dermatol. Venereol. Leprol. https://doi.org/10.4103/ijdvl.IJDVL_1143_16.

Gou, J.Y., Felippes, F.F., Liu, C.J., Weigel, D., Wang, J.W., 2011. Negative Regulation of Anthocyanin Biosynthesis in Arabidopsis by a miR156-Targeted SPL Transcription Factor. Plant Cell. https://doi.org/10.1105/tpc.111.084525.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome

assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. https:// doi.org / 10.1038/nbt.1883.

Gregory, P.J., Sperry, M., Wilson, A.F., 2008. Dietary supplements for osteoarthritis. Am. Fam. Physician 77(2), 177-84.

Grimmig, B., Matern, U., 1997. Structure of the parsley caffeoyl-CoA O-methyltransferase gene, harbouring a novel elicitor responsive cis-acting element. Plant Mol. Biol. https://doi.org/10.1023/A:1005780529457.

Grynkiewicz, G., lifirski, P., 2012. Curcumin and curcuminoids in quest for medicinal status. Acta Biochim. Pol. https://doi.org/10.18388/abp.2012_2139.

Gu, Z., Zhu, J., Hao, Q., Yuan, Y.W., Duan, Y.W., Men, S., Wang, Q., Hou, Q., Liu, Z.A., Shu, Q., Wang, L., 2019. A Novel R2R3-MYB Transcription Factor Contributes to Petal Blotch Formation by Regulating Organ-Specific Expression of PsCHS in Tree Peony (*Paeonia suffruticosa*). Plant Cell Physiol. https://doi.org/10.1093/pcp/pcy232.

Guddadarangavvanahally, J., Lingamullu Jagan, M.R., Kunnumpurath K., S., 2002. Improved HPLC Method for the Determination of Curcumin, Demethoxycurcumin, and Bisdemethoxycurcumin. J. Agric. Food Chem. 10.1021/jf025506a.

Guo, N., Cheng, F., Wu, J., Liu, B., Zheng, S., Liang, J., Wang, X., 2014. Anthocyanin biosynthetic genes in *Brassica rapa*. BMC Genomics 15, 1–11. https://doi.org/10.1186/1471-2164-15-426.

Guo, N., Han, S., Zong, M., Wang, G., Zheng, S., Liu, F., 2019. Identification and differential expression analysis of anthocyanin biosynthetic genes in leaf color variants of ornamental kale. BMC Genomics. https://doi.org/10.1186/s12864-019-5910-z.

Gupta, O.P., Karkute, S.G., Banerjee, S., Meena, N.L., Dahuja, A., 2017. Contemporary understanding of miRNA-based regulation of secondary metabolites biosynthesis in plants. Front. Plant Sci. https://doi.org/10.3389/fpls.2017.00374.

Gupta, S.C., Sung, B., Kim, J.H., Prasad, S., Li, S., Aggarwal, B.B., 2013. Multitargeting by turmeric, the golden spice: From kitchen to clinic. Mol. Nutr. Food Res. https://doi.org/10.1002/mnfr.201100741.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. https://doi.org/10.1038/13690.

Haas, M., Sprenger, H., Zuther, E., Peters, R., Seddig, S., Walther, D., Kopka, J., Hincha, D.K., Köhl, K.I., 2020. Can Metabolite- and Transcript-Based Selection for Drought Tolerance in *Solanum tuberosum* Replace Selection on Yield in Arid Environments? Front. Plant Sci. https://doi.org/10.3389/fpls.2020.01071.

Hack, C.J., 2004. Integrated transcriptome and proteome data: the challenges ahead. Brief. Funct. Genomic. Proteomic. 3, 212–219. https://doi.org/10.1093/BFGP/3.3.212.

Hadi, S., Artanti, A.N., Rinanto, Y., Wahyuni, D.S.C., 2018. Curcuminoid content of *Curcuma longa* L. and *Curcuma xanthorrhiza* rhizome based on drying method with NMR and HPLC-UVD, in: IOP Conference Series: Materials Science and Engineering. https://doi.org/10.1088/1757-899X/349/1/012058.

Hahlbrock, K., Scheel, D., 1989. Physiology and Molecular Biology of Phenylpropanoid Metabolism. Annu. Rev. Plant Physiol. Plant Mol. Biol. https://doi.org/ 10.1146 /annurev.pp.40.060189.002023.

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. Can. J. Bot. https://doi.org/10.1139/B07-098.

Hanson, K.R., Havir, E.A., 1979. An Introduction to the Enzymology of Phenylpropanoid Biosynthesis, in: Biochemistry of Plant Phenolics. https://doi.org/10.1007/978-1-4684-3372-2_4.

Hao, C., Xia, Z., Fan, R., Tan, L., Hu, L., Wu, B., Wu, H., 2016. De novo transcriptome sequencing of black pepper (*Piper nigrum* L.) and an analysis of genes involved in phenylpropanoid metabolism in response to *Phytophthora capsici*. BMC Genomics. https://doi.org/10.1186/s12864-016-3155-7.

Harvey, S.E., Lyu, J., Cheng, C., 2016. Methods for Characterization of Alternative RNA Splicing. Methods Mol. Biol. 2372, 209–222. https://doi.org/10.1007/978-1-0716-1697-0_19.

Hashemitabar, M., Kolahi, M., Tabandeh, M.R., Jonoubi, P., Majd, A., 2014. cDNA cloning, phylogenic analysis and gene expression pattern of phenylalanine ammonia-lyase in sugarcane (*Saccharum officinarum* L.). Brazilian Arch. Biol. Technol. 57, 456–465. https://doi.org/10.1590/S1516-8913201402061.

He, M., Yao, Y., Li, Yanni, Yang, M., Li, Yu, Wu, B., Yu, D., 2019. Comprehensive transcriptome analysis reveals genes potentially involved in isoflavone biosynthesis in *Pueraria thomsonii* Benth. PLoS One. https:// doi.org/ 10.1371/journal.pone.0217593.

He, S.M., Liang, Y.L., Cong, K., Chen, G., Zhao, X., Zhao, Q.M., Zhang, J.J., Wang, X., Dong, Y., Yang, J.L., Zhang, G.H., Qian, Z.L., Fan, W., Yang, S.C., 2018. Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in coptis species. Front. Plant Sci. https://doi.org/10.3389/fpls.2018.00731.

Henrotin, Y., Clutterbuck, A.L., Allaway, D., Lodwig, E.M., Harris, P., Mathy-Hartert, M., Shakibaei, M., Mobasheri, A., 2010. Biological actions of curcumin on articular chondrocytes. Osteoarthr. Cartil. https://doi.org/ 10.1016/ j. joca. 2009.10.002.

Henrotin, Y., Priem, F., Mobasheri, A., 2013. Curcumin: A new paradigm and therapeutic opportunity for the treatment of osteoarthritis: Curcumin for osteoarthritis management. Springerplus. https://doi.org/10.1186/2193-1801-2-56.

Her, C., Venier-Julienne, M.-C., Roger, E., 2018. Improvement of Curcumin Bioavailability for Medical Applications. Med. Aromat. Plants. https://doi.org/10.4172/2167-0412.1000326.

Hermes, J.D., Weiss, P.M., Cleland, W.W., 1985. Use of Nitrogen-15 and Deuterium Isotope Effects to Determine the Chemical Mechanism of Phenylalanine Ammonia-lyase. Biochemistry. https://doi.org/10.1021/bi00333a023.

Hewlings, S.J., Kalman, D.S., 2017. Curcumin: A review of its effects on human health. Foods. https://doi.org/10.3390/foods6100092.

Holt, R.A., Jones, S.J.M., 2008. The new paradigm of flow cell sequencing. Genome Res. https://doi.org/10.1101/gr.073262.107.

Hong, J.C., 2016. General Aspects of Plant Transcription Factor Families, in: Plant Transcription Factors: Evolutionary, Structural and Functional Aspects. https://doi.org/10.1016/B978-0-12-800854-6.00003-8.

Hong, Y., Li, M., Dai, S., 2019. ITRAQ-based protein profiling provides insights into the mechanism of light-induced anthocyanin biosynthesis in chrysanthemum (*Chrysanthemum morifolium*). Genes (Basel). https://doi.org/10.3390/ genes 10121024.

Hou, X., Shao, F., Ma, Y., Lu, S., 2013. The phenylalanine ammonia-lyase gene family in *Salvia miltiorrhiza*: Genome-wide characterization, molecular cloning and expression analysis. Mol. Biol. Rep. https://doi.org/10.1007/s11033-013-2517-3.

Howlader, J., Robin, A.H.K., Natarajan, S., Biswas, M.K., Sumi, K.R., Song, C.Y., Park, J. –I, Nou, I. –S, 2020. Transcriptome Analysis by RNA–Seq Reveals Genes Related to Plant Height in Two Sets of Parent-hybrid Combinations in Easter lily (*Lilium longiflorum*). Sci. Reports 2020 101 10, 1–15. https://doi.org/ 10.1038/ s41598-020-65909-x.

Hrazdina, G., Wagner, G.J., 1985. Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. Arch. Biochem. Biophys. https://doi.org/10.1016/0003-9861(85)90257-7.

Hu, L., Hao, C., Fan, R., Wu, B., Tan, L., Wu, H., 2015. De novo assembly and characterization of fruit transcriptome in black pepper (*Piper nigrum*). PLoS One. https://doi.org/10.1371/journal.pone.0129822

Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., Rhee, S.Y., 2001. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. https://doi.org/10.1093/nar/29.1.102.

Huang, T.S., Anzellotti, D., Dedaldechamp, F., Ibrahim, R.K., 2004. Partial purification, kinetic analysis, and amino acid sequence information of a flavonol 3-O-methyltransferase from *Serratula tinctoria*. Plant Physiol. https://doi.org/10.1104/ pp.103.036442.

Hugueney, P., Provenzano, S., Verriès, C., Ferrandino, A., Meudec, E., Batelli, G., Merdinoglu, D., Cheynier, V., Schubert, A., Ageorges, A., 2009. A novel cation-dependent o-methyltransferase involved in anthocyanin methylation in grapevine. Plant Physiol. https://doi.org/10.1104/pp.109.140376.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., Sigrist, C.J.A., 2006. The PROSITE database. Nucleic Acids Res. https://doi.org/10.1093/nar/gkj063.

Hussain, M., Ahmad Khan, S., Noureen, N., Fatima, S., Ul Ane, N., Abbas, Z., 2015. Yield Performance of Turmeric Varieties Intercropped with Mulberry Plantations. J. Agric. Environ. Sci.

Ibdah, M., Zhang, X.H., Schmidt, J., Vogt, T., 2003. A Novel Mg2+-dependent O-Methyltransferase in the Phenylpropanoid Metabolism of Mesembryanthemum crystallinum. J. Biol. Chem. https://doi.org/10.1074/jbc.M304932200.

Indira Priyadarsini, K., 2013. Chemical and Structural Features Influencing the Biological Activity of Curcumin. Curr. Pharm. Des. https://doi.org/10.2174/138161213805289228.

Inoue, K., Sewalt, V.J.H., Ballance, G.M., Ni, W., Stürzer, C., Dixon, R.A., 1998. Developmental expression and substrate specificities of alfalfa caffeic acid 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase in relation to lignification. Plant Physiol. https://doi.org/10.1104/pp.117.3.761.

Islam, M.S., Jalaluddin, M., Garner, J.O., Yoshimoto, M., Yamakawa, O., 2005. Artificial shading and temperature influence on anthocyanin compositions in sweetpotato leaves. HortScience 40, 176–180. https://doi.org/10.21273/HORTSCI.40.1.176.

Jaakola, L., Poole, M., Jones, M.O., Kämäräinen-Karppinen, T., Koskimäki, J.J., Hohtola, A., Häggman, H., Fraser, P.D., Manning, K., King, G.J., Thomson, H., Seymour, G.B., 2010. A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. Plant Physiol. https://doi.org/ 10.1104/pp.110.158279.

Jahnen, W., Hahlbrock, K., 1988. Differential regulation and tissue-specific distribution of enzymes of phenylpropanoid pathways in developing parsley seedlings. Planta. https://doi.org/10.1007/BF00958957.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F., 2002. bZIP transcription factors in Arabidopsis. Trends Plant Sci. https://doi.org/10.1016/S1360-1385(01)02223-3.

Jakubczyk, K., Druga, A., Katarzyna, J., Skonieczna ydecka, K., 2020. Antioxidant potential of curcumin a meta-analysis of randomized clinical trials. Antioxidants. https://doi.org/10.3390/antiox9111092.

Jamwal, R., 2018. Bioavailable curcumin formulations: A review of pharmacokinetic studies in healthy volunteers. J. Integr. Med. https://doi.org/10.1016/ j. joim.2018.07.001.

Janas, K.M., 1993. The control of L-phenylalanine ammonia-lyase activity by phosphonate and aminooxy analogues of phenylalanine. Acta Biochim. Pol. https://doi.org/10.18388/abp.1993_4783.

Jarret, R.L., Barboza, G.E., Batista, F.R. da C., Berke, T., Chou, Y.Y., Hulse-Kemp, A., Ochoa-Alejo, N., Tripodi, P., Veres, A., Garcia, C.C., Csillery, G., Huang, Y.K., Kiss, E., Kovacs, Z., Kondrak, M., Arce-Rodriguez, M.L., Scaldaferro, M.A., Szoke, A., 2019. Capsicum-an abbreviated compendium. J. Am. Soc. Hortic. Sci. https://doi.org/10.21273/JASHS04446-18.

Jayaprakasha, G.K., Rao, L.J.M., Sakariah, K.K., 2002. Improved HPLC method for the determination of curcumin, demethoxycurcumin, and bisdemethoxycurcumin. J. Agric. Food Chem. https://doi.org/10.1021/jf025506a.

Jiang, C., Zhang, H., Ren, J., Dong, J., Zhao, X., Wang, X., Wang, J., Zhong, C., Zhao, S., Liu, X., Gao, S., Yu, H., 2020. Comparative transcriptome-based mining and expression profiling of transcription factors related to cold tolerance in peanut. Int. J. Mol. Sci. https://doi.org/10.3390/ijms21061921.

Jiang, Y., Liao, Q., Zou, Y., Liu, Y., Lan, J., 2017. Transcriptome analysis reveals the genetic basis underlying the biosynthesis of volatile oil, gingerols, and diarylheptanoids in ginger (*Zingiber officinale* Rosc.). Bot. Stud. https://doi.org/10.1186/s40529-017-0195-5.

Jiang, Y., Xia, N., Li, X., Shen, W., Liang, L., Wang, C., Wang, R., Peng, F., Xia, B., 2011. Molecular cloning and characterization of a phenylalanine ammonia-lyase

gene (LrPAL) from *Lycoris radiata*. Mol. Biol. Rep. https://doi.org/ 10.1007/s11033-010-0314-9.

Jin, Q., Yao, Y., Cai, Y., Lin, Y., 2013. Molecular Cloning and Sequence Analysis of a Phenylalanine Ammonia-Lyase Gene from Dendrobium. PLoS One 8. https://doi.org/10.1371/journal.pone.0062352.

Jin, Z., Kwon, M., Lee, A.R., Ro, D.K., Wungsintaweekul, J., Kim, S.U., 2018. Molecular cloning and functional characterization of three terpene synthases from unripe fruit of black pepper (*Piper nigrum*). Arch. Biochem. Biophys. https://doi.org/ 10.1016/j.abb.2017.12.011.

Jones, D.H., 1984. Phenylalanine ammonia-lyase: Regulation of its induction, and its role in plant development. Phytochemistry. https://doi.org/10.1016/S0031-9422(00)80465-3.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu031.

Joos, H. J, Hahlbrock, K., 1992. Phenylalanine ammonia lyase in potato (*Solanum tuberosum* L.): Genomic complexity, structural comparison of two selected genes and modes of expression. Eur. J. Biochem. https://doi.org/10.1111/j.1432-1033.1992.tb16675-x.

Jorrin-Novo, J. V., 2014. Plant proteomics methods and protocols. Methods Mol. Biol. https://doi.org/10.1007/978-1-62703-631-3_1.

Joshi, C.P., Chiang, V.L., 1998. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. Plant Mol. Biol. https://doi.org/ 10.1023/ A:1006035210889.

Joshi, R., Pareek, A., Singla-Pareek, S.L., 2015. Plant Metallothioneins: Classification, Distribution, Function, and Regulation. Plant Met. Interact. Emerg. Remediat. Tech. 239–261. https://doi.org/10.1016/B978-0-12-803158-2.00009-6.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. https://doi.org/10.1093/nar/gkv1070.

Kang, W.H., Sim, Y.M., Koo, N., Nam, J.Y., Lee, J., Kim, N., Jang, H., Kim, Y.M., Yeom, S.I., 2020. Transcriptome profiling of abiotic responses to heat, cold, salt, and osmotic stress of *Capsicum annuum* L. Sci. Data. https://doi.org/10.1038/s41597-020-0352-7.

Katiyar, A., Smita, S., Lenka, S.K., Rajwanshi, R., Chinnusamy, V., Bansal, K.C., 2012. Genome-wide classification and expression analysis of MYB transcription factor

families in rice and Arabidopsis. BMC Genomics. https://doi.org/10.1186/1471-2164-13-544.

Katsuyama, Y., Kita, T., Funa, N., Horinouchi, S., 2009a. Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. J. Biol. Chem. https://doi.org/10.1074/jbc.M900070200.

Katsuyama, Y., Kita, T., Funa, N., Horinouchi, S., 2009a. Curcuminoid Biosynthesis by Two Type III Polyketide Synthases in the Herb *Curcuma longa* S. J. Biol. Chem. 284, 11160–11170. https://doi.org/10.1074/jbc.M900070200.

Katsuyama, Y., Kita, T., Funa, N., Horinouchi, S., 2009b. Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. J. Biol. Chem. https://doi.org/10.1074/jbc.M900070200.

Katsuyama, Y., Kita, T., Horinouchi, S., 2009b. Identification and characterization of multiple curcumin synthases from the herb *Curcuma longa*. FEBS Lett. https://doi.org/10.1016/j.febslet.2009.07.029.

Khew, C.Y., Harikrishna, J.A., Wee, W.Y., Lau, E.T., Hwang, S.S., 2020. Transcriptional Sequencing and Gene Expression Analysis of Various Genes in Fruit Development of Three Different Black Pepper (*Piper nigrum* L.) Varieties. Int. J. Genomics. https://doi.org/10.1155/2020/1540915.

Khew, C.Y., Harikrishna, J.A., Wee, W.Y., Lau, E.T., Hwang, S.S., 2020. Transcriptional Sequencing and Gene Expression Analysis of Various Genes in Fruit Development of Three Different Black Pepper (*Piper nigrum* L.) Varieties. Int. J. Genomics. https://doi.org/10.1155/2020/1540915.

Kilner, J., Zhu, L., Ow, S.Y., Evans, C., Corfe, B.M., 2011. Assessing the loss of information through application of the 'two-hit rule' in iTRAQ datasets. J. Integr. OMICS 1, 124–134. https://doi.org/10.5584/JIOMICS.V1I1.53.

Kim, S., 2016. Getting the most out of PubChem for virtual screening. Expert Opin. Drug Discov. https://doi.org/10.1080/17460441.2016.1216967.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H., 2016. PubChem substance and compound databases. Nucleic Acids Res. https://doi.org/10.1093/nar/gkv951.

Kim, S.H., Lee, J.R., Hong, S.T., Yoo, Y.K., An, G., Kim, S.R., 2003. Molecular cloning and analysis of anthocyanin biosynthesis genes preferentially expressed in apple skin. Plant Sci. https://doi.org/10.1016/S0168-9452(03)00201-2.

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. Biosci. Biotechnol. Biochem. https://doi.org/10.1271/bbb.80075.

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. Biosci. Biotechnol. Biochem. https://doi.org/10.1271/bbb.80075.

Kodama, M., Brinch-Pedersen, H., Sharma, S., Holme, I.B., Joernsgaard, B., Dzhanfezova, T., Amby, D.B., Vieira, F.G., Liu, S., Gilbert, M.T.P., 2018. Identification of transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. BMC Genomics. https://doi.org/10.1186/s12864-018-5135-6.

Koirala, N., Thuan, N.H., Ghimire, G.P., Thang, D. Van, Sohng, J.K., 2016. Methylation of flavonoids: Chemical structures, bioactivities, progress and perspectives for biotechnological production. Enzyme Microb. Technol. 86, 103–116. https://doi.org/10.1016/J.ENZMICTEC.2016.02.003.

Korir, N.K., Han, J., Shangguan, L., Wang, C., Kayesh, E., Zhang, Y., Fang, J., 2013. Plant variety and cultivar identification: Advances and prospects. Crit. Rev. Biotechnol. https://doi.org/10.3109/07388551.2012.675314.

Kosuge, T., Conn, E.E., 1962. The Metabolism of Aromatic Compounds in Higher Plants. J. Biol. Chem. https://doi.org/10.1016/s0021-9258(19)83757-2.

Koukol, J., Conn, E.E., 1961. The metabolism of aromatic compounds in higher plants. IV. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare*. J. Biol. Chem. 236, 2692–2698.

Kremsner, P.G., Winkler, S., Brandts, C., Neifer, S., Bienzel, U., Graninger, W., 1994. Clindamycin in combination with chloroquine or quinine is an effective therapy for uncomplicated *Plasmodium falciparum* malaria in children from gabon. J. Infect. Dis. https://doi.org/10.1093/infdis/169.2.467.

Kumar Ayer, D., Modha, K., Parekh, V., Patel, R., Vadodariya, G., Ramtekey, V., Bhuriya, A., 2020. Associating gene expressions with curcuminoid biosynthesis in turmeric. https://doi.org/10.1186/s43141-020-00101-2.

Kumar, A., Ellis, B.E., 2001. The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression, and evolution. Plant Physiol. https://doi.org/ 10.1104/ pp.127.1.230.

Lal, J., 2012. Turmeric, Curcumin and Our Life: A Review. Bull. Environ. Pharmacol. Life Sci.

Lamb, C.J., Merritt, T.K., Butt, V.S., 1979. Synthesis and removal of phenylalanine ammonia-lyase activity in illuminated discs of potato tuber parenchyme. BBA - Gen. Subj. https://doi.org/10.1016/0304-4165(79)90384-2.

Lampe, V., 1918. Synthese von Curcumin. Berichte der Dtsch. Chem. Gesellschaft 51, 1347–1355. https://doi.org/https://doi.org/10.1002/cber.19180510223.

Langmead, B., 2010. Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinforma. https://doi.org/10.1002/0471250953.bi1107s32.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. https://doi.org/10.1107/s0021889892009944.

Le Roy, J., Huss, B., Creach, A., Hawkins, S., Neutelings, G., 2016. Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.00735.

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. https://doi.org/10.1007/978-981-10-1085-9_1.

Levac, D., Murata, J., Kim, W.S., De Luca, V., 2008. Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O- methyltransferase. Plant J. https:// doi.org/ 10.1111/j.1365-313X.2007.03337-x.

Li, H., Sureda, A., Devkota, H.P., Pittalà, V., Barreca, D., Silva, A.S., Tewari, D., Xu, S., Nabavi, S.M., 2020. Curcumin, the golden spice in treating cardiovascular diseases. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2019.01.010.

Li, Jiewen, Zhao, W., Li, Jian, 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, L., Popko, J.L., Umezawa, T., Chiang, V.L., 2000. 5-Hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms. J. Biol. Chem. https://doi.org/10.1074/jbc.275.9.6537.

Li, L., Popko, J.L., Zhang, X.H., Osakabe, K., Tsai, C.J., Joshi, C.P., Chiang, V.L., 1997. A novel multifunctional O-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.94.10.5461.

Li, R., Reddy, V.A., Jin, J., Rajan, C., Wang, Q., Yue, G., Lim, C.H., Chua, N.H., Ye, J., Sarojam, R., 2017. Comparative transcriptome analysis of oil palm flowers reveals an EAR-motif-containing R2R3-MYB that modulates phenylpropene biosynthesis. BMC Plant Biol. https://doi.org/10.1186/s12870-017-1174-4.

Li, S., 2011. Chemical Composition and Product Quality Control of Turmeric (*Curcuma longa* L.). Pharm. Crop. https://doi.org/10.2174/2210290601102010028.

Li, S., 2014. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. Plant Signal. Behav. https://doi.org/ 10. 4161/psb.27522.

Li, X.-S., Yang, H.-L., Zhang, D.-Y., Zhang, Y.-M., Wood, A.J., 2012. Reference Gene Selection in the Desert Plant *Eremosparton songoricum*. Int. J. Mol. Sci. https://doi.org/10.3390/ijms13066944.

Li, Y., Shan, X., Zhou, L., Gao, R., Yang, S., Wang, S., Wang, L., Gao, X., 2019. The R2R3-MYB factor FhMYB5 from *Freesia hybrida* contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis. Front. Plant Sci. https://doi.org/10.3389/fpls.2018.01935.

Li, Y.X., Pan, Y.G., He, F.P., Yuan, M.Q., Li, S. Bin, 2016. Pathway analysis and metabolites identification by Metabolomics of etiolation substrate from fresh-cut Chinese water chestnut (*Eleocharis tuberosa*). Molecules. https://doi.org/ 10.3390/ molecules21121648.

Liang, M., Raley, C., Zheng, X., Kutty, G., Gogineni, E., Sherman, B.T., Sun, Q., Chen, X., Skelly, T., Jones, K., Stephens, R., Zhou, B., Lau, W., Johnson, C., Imamichi, T., Jiang, M., Dewar, R., Lempicki, R.A., Tran, B., Kovacs, J.A., Huang, D.W., 2016. Distinguishing highly similar gene isoforms with a clustering-based bioinformatics analysis of PacBio single-molecule long reads. BioData Min. 9, 1–13. https://doi.org/10.1186/s13040-016-0090-8.

Liang, P., Pardee, A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science (80-.). https:// doi. org/ 10. 1126/science.1354393.

Liang, X.W., Dron, M., Cramer, C.L., Dixon, R.A., Lamb, C.J., 1989. Differential regulation of phenylalanine ammonia-lyase genes during plant development and by environmental cues. J. Biol. Chem. https://doi.org/10.1016/s0021-9258(18)71704-3.

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 Environ. https://doi.org/10.1111/j.1365-3040.2007.01748-x.

Lin, J.K., Chen, Y.C., Huang, Y.T., Lin-Shiau, S.Y., 1997. Suppression of protein kinase C and nuclear oncogene expression as possible molecular mechanisms of cancer chemoprevention by apigenin and curcumin. J. Cell. Biochem. https://doi.org/10.1002/(SICI)1097-4644(1997)28/29+<39

Lin, L.I., Ke, Y.F., Ko, Y.C., Lin, J.K., 1998. Curcumin inhibits SK-Hep-1 hepatocellular carcinoma cell invasion *in vitro* and suppresses matrix metalloproteinase-9 secretion. Oncology. https://doi.org/10.1159/000011876.

Lin, Q., Zhong, Q., Zhang, Z., 2019. Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the pink-white and red fruits of Chinese bayberry (*Morella rubra*). Sci. Hortic. (Amsterdam). https://doi.org/ 10. 1016/j.scienta.2019.02.061.

Lister, C.E., Lancaster, J.E., Walker, J.R.L., 1996. Developmental changes in enzymes of flavonoid biosynthesis in the skins of red and green apple cultivars. J. Sci. Food Agric. https://doi.org/10.1002/(SICI) 1097- 0010(199607) 71:3<313: AID-JSFA586> 3.0.CO;2-N.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., Ecker, J.R., 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell. https://doi.org/10.1016/j.cell.2008.03.029.

Liu, C., Long, J., Zhu, K., Liu, L., Yang, W., Zhang, H., Li, L., Xu, Q., Deng, X., 2016. Characterization of a Citrus R2R3-MYB Transcription Factor that Regulates the Flavonol and Hydroxycinnamic Acid Biosynthesis. Sci. Rep. https://doi.org/10.1038/srep25352.

Liu, J., Yuan, Y., Wang, Y., Jiang, C., Chen, T., Zhu, F., Zhao, Y., Zhou, J., Huang, L., 2017. Regulation of fatty acid and flavonoid biosynthesis by miRNAs in: *Lonicera japonica*. RSC Adv. https://doi.org/10.1039/c7ra05800d.

Liu, P., Wang, Y., Meng, J., Zhang, X., Zhou, J., Han, M., Yang, C., Gan, L., Li, H., 2019. Transcriptome sequencing and expression analysis of genes related to anthocyanin biosynthesis in leaves of Malus "Profusion" infected by Japanese apple rust. Forests. https://doi.org/10.3390/f10080665.

Liu, S., Li, W., Wu, Y., Chen, C., Lei, J., 2013. De Novo Transcriptome Assembly in Chili Pepper (*Capsicum frutescens*) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. PLoS One. https://doi.org/10.1371/journal.pone.0048156.

Liu, S.R., Zhou, J.J., Hu, C.G., Wei, C.L., Zhang, J.Z., 2017a. MicroRNA-mediated gene silencing in plant defense and viral counter-defense. Front. Microbiol. https://doi.org/10.3389/fmicb.2017.01801.

Liu, W., Wang, Y., Yu, L., Jiang, H., Guo, Z., Xu, H., Jiang, S., Fang, H., Zhang, J., Su, M., Zhang, Z., Chen, Xiaoliu, Chen, Xuesen, Wang, N., 2019. MdWRKY11 Participates in Anthocyanin Accumulation in Red-Fleshed Apples by Affecting MYB Transcription Factors and the Photoresponse Factor MdHY5. J. Agric. Food Chem. https://doi.org/10.1021/acs.jafc.9b02920.

Liu, X., Wang, Y., Chen, Y., Xu, S., Gong, Q., Zhao, C., Cao, J., Sun, C., 2020a. Characterization of a Flavonoid 3'/5'/7-O-Methyltransferase from *Citrus reticulata* and Evaluation of the *in vitro* Cytotoxicity of Its Methylated Products. Molecules. https://doi.org/10.3390/molecules25040858.

Liu, X., Zhao, C., Gong, Q., Wang, Y., Cao, J., Li, X., Grierson, D., Sun, C., 2020b. Characterization of a caffeoyl-CoA O-methyltransferase-like enzyme involved in biosynthesis of polymethoxylated flavones in *Citrus reticulata*. J. Exp. Bot. https://doi.org/10.1093/jxb/eraa083.

Liu, Y., Chaturvedi, P., Fu, J., Cai, Q., Weckwerth, W., Yang, P., 2016a. Induction and quantitative proteomic analysis of cell dedifferentiation during callus formation of lotus (*Nelumbo nucifera* Gaertn.spp. baijianlian). J. Proteomics. https://doi.org/10.1016/j.jprot.2015.10.010.

Liu, Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., Yang, H., Wang, J., Espley, R. V., Zhang, J., Wang, D., Allan, A.C., 2015. Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. PLoS One. https://doi.org/10.1371/journal.pone.0129148.

Llave, C., Kasschau, K.D., Rector, M.A., Carrington, J.C., 2002. Endogenous and silencing-associated small RNAs in plants. Plant Cell. https://doi.org/ 10.1105/tpc.003210.

Lloyd, A., Brockman, A., Aguirre, L., Campbell, A., Bean, A., Cantero, A., Gonzalez, A., 2017. Advances in the MYB-bHLH-WD Repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. Plant Cell Physiol. https://doi.org/10.1093/pcp/pcx075.

Logemann, E., Parniske, M., Hahlbrock, K., 1995. Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.92.13.5905.

Lois, R., Dietrich, A., Hahlbrock, K., Schulz, W., 1989. A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. EMBO J. https://doi.org/10.1002/j.1460-2075.1989.tb03554-x.

Louie, G. V., Bowman, M.E., Tu, Y., Mouradov, A., Spangenberg, G., Noel, J.P., 2010. Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. Plant Cell. https://doi.org/10.1105/tpc.110.077578.

Lu, J., Zhang, Q., Lang, L., Jiang, C., Wang, X., Sun, H., 2021. Integrated metabolome and transcriptome analysis of the anthocyanin biosynthetic pathway in relation to color mutation in miniature roses. BMC Plant Biol. https://doi.org/ 10.1186/ s12870-021-03063-w.

Lu, Y., Zhang, M., Meng, X., Wan, H., Zhang, J., Tian, J., Hao, S., Jin, K., Yao, Y., 2015. Photoperiod and shading regulate coloration and anthocyanin accumulation in the leaves of Malus crabapples. Plant Cell. Tissue Organ Cult. 121, 619-632. https://doi.org/10.1007/S11240-015-0733-3.

Luo, R., and Zhao, H. 2012. Protein quantitation using iTRAQ: Review on the sources of variations and analysis of nonrandom missingness. Stat. Interface, *5*(1), 99.

Luo, R., Zhao, H., 2012. Protein quantitation using iTRAQ: Review on the sources of variations and analysis of nonrandom missingness. Stat. Interface, 5, 99. https://doi.org/10.4310/sii.2012.v5.n1.a9.

Luo, Y., Zhang, X., Luo, Z., Zhang, Q., Liu, J., 2015. Identification and characterization of microRNAs from Chinese pollination constant non-astringent persimmon using high-throughput sequencing. BMC Plant Biol. https://doi.org/10.1186/s12870-014-0400-6.

Ma, D., Reichelt, M., Yoshida, K., Gershenzon, J., Constabel, C.P., 2018. Two R2R3-MYB proteins are broad repressors of flavonoid and phenylpropanoid metabolism in poplar. Plant J. https://doi.org/10.1111/tpj.14081.

Ma, H., Pooler, M., Griesbach, R., 2009. Anthocyanin regulatory/structural gene expression in Phalaenopsis. J. Am. Soc. Hortic. Sci. https://doi.org/ 10.21273/ jashs.134.1.88.

Ma, Q.H., Luo, H.R., 2015. Biochemical characterization of caffeoyl coenzyme A 3-O-methyltransferase from wheat. Planta. https://doi.org/10.1007/s00425-015-2295-3.

Ma, Q.H., Luo, H.R., 2015. Biochemical characterization of caffeoyl coenzyme A 3-O-methyltransferase from wheat. Planta. https://doi.org/10.1007/s00425-015-2295-3.

Ma, Q.H., Xu, Y., 2008. Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis. Biochimie. https://doi.org/ 10. 1016/j.biochi.2007.09.016.

Ma, W.L., Wu, M., Wu, Y., Ren, Z.M., Zhong, Y., 2013. Cloning and characterisation of a phenylalanine ammonia-lyase gene from *Rhus chinensis*. Plant Cell Rep. 32, 1179-1190. https://doi.org/10.1007/s00299-013-1413-6.

Mahesh, V., Rakotomalala, J.J., Le Gal, L., Vigne, H., De Kochko, A., Hamon, S., Noirot, M., Campa, C., 2006. Isolation and genetic mapping of a *Coffea canephora* phenylalanine ammonia-lyase gene (CcPAL1) and its involvement in the accumulation of caffeoyl quinic acids. Plant Cell Rep. https://doi.org/ 10.1007/s00299-006-0152-3.

Maier, T., Güell, M., Serrano, L., 2009. Correlation of mRNA and protein in complex biological samples. FEBS Lett. https://doi.org/10.1016/j.febslet.2009.10.036.

Malacarne, G., Coller, E., Czemmel, S., Vrhovsek, U., Engelen, K., Goremykin, V., ... & Moser, C. (2016). The grapevine VvibZIPC22 transcription factor is involved in the regulation of flavonoid biosynthesis. *Journal of experimental botany*, *67*(11), 3509-3522.

Malhotra, K., Foltz, L., Mahoney, W.C., Schueler, P.A., 1998. Interaction and effect of annealing temperature on primers used in differential display RT-PCR. Nucleic Acids Res. https://doi.org/10.1093/nar/26.3.854.

Manzoni, C., Kia, D.A., Vandrovcova, J., Hardy, J., Wood, N.W., Lewis, P.A., Ferrari, R., 2018. Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. Brief. Bioinform. https://doi.org/ 10.1093/BIB/BBW114.

Marcela, V. H., Gerardo, V. M., Agustín, A. R. C., Antonio, G. M. M., Oscar, R., Diego, C. P., & Cruz-Hernández, A. 2019. MicroRNAs associated with secondary metabolites production. Plant Physiological Aspects of Phenolic Compounds.https://doi.org/10.5772/intechopen.83804.

Martin, C., Paz-Ares, J., 1997. MYB transcription factors in plants. Trends Genet. https://doi.org/10.1016/S0168-9525(96)10049-4.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. https://doi.org/10.14806/ej.17.1.200.

Martínez, O., Arce-Rodríguez, M.L., Hernández-Godínez, F., Escoto-Sandoval, C., Cervantes-Hernández, F., Hayano-Kanashiro, C., Ordaz-Ortiz, J.J., Humberto Reyes-Valdés, M., Razo-Mendivil, F.G., Garcés-Claver, A., Ochoa-Alejo, N., 2021. Transcriptome analyses throughout chili pepper fruit development reveal novel insights into the domestication process. Plants. https://doi.org/10.3390/plants10030585.

Martz, F., Maury, S., Pinçon, G., Legrand, M., 1998. cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-O-methyltransferase, a lignin biosynthetic enzyme. Plant Mol. Biol. https://doi.org/10.1023/A:1005969825070.

Maury, S., Geoffroy, P., Legrand, M., 1999. Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate spec. Plant Physiol. 121, 215–223. https://doi.org/10.1104/pp.121.1.215.

Mehrtens, F., Kranz, H., Bednarek, P., Weisshaar, B., 2005. The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiol. https://doi.org/10.1104/pp.104.058032.

Melo, C.A., Melo, S.A., 2014. MicroRNA biogenesis: Dicing assay. Methods Mol. Biol. https://doi.org/10.1007/978-1-4939-1062-5_20.

Menon, V.P., Sudheer, A.R., 2007. Antioxidant and anti-inflammatory properties of curcumin. Adv. Exp. Med. Biol. https://doi.org/10.1007/978-0-387-46401-5_3.

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., Herdewijn, P., Devreese, B., Van Beeumen, J., Marita, J.M., Ralph, J., Chen, C.,

Burggraeve, B., Van Montagu, M., Messens, E., Boerjan, W., 2000. Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. J. Biol. Chem. https://doi.org/10.1074/jbc.M006915200.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H., Thomas, P.D., 2005. The PANTHER database of protein families, sub-families, functions and pathways. Nucleic Acids Res. https://doi.org/10.1093/nar/gki078

Miao, L., Shou, S., Zhu, Z., Jiang, F., Zai, W., Yang, Y., 2008. Isolation of a novel tomato caffeoyl CoA 3-O-methyltransferase gene following infection with the bacterium *Ralstonia solanacearum*. J. Phytopathol. https://doi.org/10.1111/j.1439-0434. 2008.01406.

Miłobędzka, J., v. Kostanecki, S., & Lampe, V., 1910. Zur kenntnis des curcumins. Berichte der deutschen chemischen Gesellschaft, 43(2), 2163-2170.

Minami, E. I, Ozeki, Y., Matsuoka, M., Koizuka, N., Tanaka, Y., 1989. Structure and some characterization of the gene for phenylalanine ammonia lyase from rice plants. Eur. J. Biochem. https://doi.org/10.1111/j.1432-1033.1989.tb15075-x.

Moglia, A., Francesco, E.F., Sergio, I., Alessandra, G., Milani, A.M., Cinzia, C., Barchi, L., Marengo, A., Cagliero, C., Rubiolo, P., Toppino, L., Rotino, L.G., Lanteri, S., Laura, B., 2020. Identification of a new R3 MYB type repressor and functional characterization of the members of the MBW transcriptional complex involved in anthocyanin biosynthesis in eggplant (*S. melongena* L.). PLoS One. https://doi.org/10.1371/journal.pone.0232986.

Morishige, T., Tamakoshi, M., Takemura, T., Sato, F., 2010. Molecular characterization of O-methyltransferases involved in isoquinoline alkaloid biosynthesis in Coptis japonica. Proc. Japan Acad. Ser. B Phys. Biol. Sci. https://doi.org/10.2183/pjab.86.757.

Muthusamy, A., 2011. A Study on Export Performance of Indian Turmeric. Indian J. Appl. Res. 3, 54–56. https://doi.org/10.15373/2249555x/apr2013/18.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (80). https://doi.org/10.1126/science.1158441.

Nair, P.P., 2013. The Agronomy and Economy of Turmeric and Ginger, The Agronomy and Economy of Turmeric and Ginger. https://doi.org/10.1016/C2011-0-07514-2.

Nakatsuka, T., Haruta, K.S., Pitaksutheepong, C., Abe, Y., Kakizaki, Y., Yamamoto, K., Shimada, N., Yamamura, S., Nishihara, M., 2008. Identification and characterization of R2R3-MYB and bHLH transcription factors regulating

anthocyanin biosynthesis in gentian flowers. Plant Cell Physiol. https://doi.org/10.1093/pcp/pcn163.

Neer, E.J., Schmidt, C.J., Nambudripad, R., Smith, T.F., 1994. The ancient regulatory-protein family of WD-repeat proteins. Nature. https://doi.org/10.1038/371297a0.

Nemesio-Gorriz, M., Blair, P.B., Dalman, K., Hammerbacher, A., Arnerup, J., Stenlid, J., Mukhtar, S.M., Elfstrand, M., 2017. Identification of Norway spruce MYB-bHLH-WDR transcription factor complex members linked to regulation of the flavonoid pathway. Front. Plant Sci. https://doi.org/10.3389/fpls.2017.00305.

Njuguna, N.M., Ongarora, D.S.B., Chibale, K., 2012. Artemisinin derivatives: A patent review (2006 present). Expert Opin. Ther. Pat. https://doi.org/ 10.1517/ 13543776.2012.724063.

Noel, J.P., Austin, M.B., Bomati, E.K., 2005. Structure-function relationships in plant phenylpropanoid biosynthesis. Curr. Opin. Plant Biol. https://doi.org/ 10.1016/ j.pbi.2005.03.013.

Noel, J.P., Dixon, R.A., Pichersky, E., Zubieta, C., Ferrer, J.L., 2003. Chapter two Structural, functional, and evolutionary basis for methylation of plant small molecules. Recent Adv. Phytochem. https://doi.org/10.1016/S0079-9920 (03) 80017-5.

Noirel, J., Evans, C., Salim, M., Mukherjee, J., Yen Ow, S., Pandhal, J., Khoa Pham, T., A. Biggs, C., C. Wright, P., 2011. Methods in Quantitative Proteomics: Setting iTRAQ on the Right Track. Curr. Proteomics. https://doi.org/ 10.2174/ 157016411794697408.

Nugroho, L.H., Verberne, M.C., Verpoorte, R., 2002. Activities of enzymes involved in the phenylpropanoid pathway in constitutively salicylic acid-producing tobacco plants. Plant Physiol. Biochem. https://doi.org/10.1016/S0981-9428(02)01437-7.

O'Neal, D., Keller, C.J., 1970. Partial purification and some properties of phenylalanine ammonia-lyase of tobacco (*Nicotiana tabacum*). Phytochemistry. https://doi.org/10.1016/S0031-9422(00)85250-4.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics. https://doi.org/10.1074/mcp.M200025-MCP200.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., Kikuchi, S., 2003. Comprehensive Analysis of NAC Family Genes in *Oryza sativa* and *Arabidopsis thaliana*. DNA Res. https:// doi.org/ 10.1093/dnares/10.6.239.

Ozsolak, F., Milos, P.M., 2011. RNA sequencing: Advances, challenges and opportunities. Nat. Rev. Genet. https://doi.org/10.1038/nrg2934.

Padhan, J.K., Kumar, P., Sood, H., Chauhan, R.S., 2016. Prospecting NGS-transcriptomes to assess regulation of miRNA-mediated secondary metabolites biosynthesis in *Swertia chirayita*, a medicinal herb of the North-Western Himalayas. Med. Plants. https://doi.org/10.5958/0975-6892.2016.00029.

Padilla-González, G.F., Frey, M., Gómez-Zeledón, J., Da Costa, F.B., Spring, O., 2019. Metabolomic and gene expression approaches reveal the developmental and environmental regulation of the secondary metabolism of yacón (*Smallanthus sonchifolius*, Asteraceae). Sci. Rep. https://doi.org/10.1038/s41598-019-49246-2.

Pan, Y., Zhao, X., Wu, X.L., Wang, Y., Tan, J., Chen, D.X., 2021. Transcriptomic and metabolomic analyses provide insights into the biosynthesis of chlorogenic acids in *Lonicera macranthoides* Hand. Mazz. PLoS One. https://doi.org/10.1371/journal.pone.0251390.

Pang, S.L., Ong, S.S., Lee, H.H., Zamri, Z., Kandasamy, K.I., Choong, C.Y., Wickneswari, R., 2014. Isolation and characterization of CCoAOMT in interspecific hybrid of *Acacia auriculiformis* x *Acacia mangium* - A key gene in lignin biosynthesis. Genet. Mol. Res. https://doi.org/10.4238/2014.

Parvathi, K., Chen, F., Guo, D., Blount, J.W., Dixon, R.A., 2001. Substrate preferences of O-methyltransferases in alfalfa suggest new pathways for 3-O-methylation of monolignols. Plant J. https://doi.org/10.1046/j.1365-313X.2001.00956.

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. https://doi.org/10.1371/journal.pone.0065622.

Pawar, K.S., Mastud, R.N., Pawar, S.K., Pawar, S.S., Bhoite, Rahul R., Bhoite, Ramesh R., Kulkarni, M. V., Deshpande, A.R., 2021. Oral Curcumin with Piperine as Adjuvant Therapy for the Treatment of COVID-19: A Randomized Clinical Trial. Front. Pharmacol. https://doi.org/10.3389/fphar.2021.669362.

Pesch, M., Schultheiß, I., Klopffleisch, K., Uhrig, J.F., Koegl, M., Clemen, C.S., Simon, R., Weidtkamp-Peters, S., Hülskamp, M., 2015. Transparent testa glabra1 and glabra1 compete for binding to glabra3 in arabidopsis. Plant Physiol. https://doi.org/10.1104/pp.15.00328.

Piccaglia, R., Marotti, M., Baldoni, G., 2002. Factors influencing anthocyanin content in red cabbage (*Brassica oleracea* var capitata L f rubra (L) Thell). J. Sci. Food Agric. 82, 1504–1509. https://doi.org/10.1002/JSFA.1226.

Piétu, G., Eveno, E., Soury-Segurens, B., Fayein, N.A., Mariage-Samson, R., Matingou, C., Leroy, E., Dechesne, C., Krieger, S., Ansorge, W., Reguigne-Arnould, I.,

Cox, D., Dehejia, A., Polymeropoulos, M.H., Devignes, M.D., Auffray, C., 1999. The Genexpress IMAGE Knowledge Base of the human muscle transcriptome: A resource of structural, functional, and positional candidate genes for muscle physiology and pathologies. Genome Res. https://doi.org/10.1101/gr.9.12.1313.

Pongrakhananon, V., Rojanasakul, Y., 2011. Anticancer Properties of Curcumin, in: Advances in Cancer Therapy. https://doi.org/10.5772/23594.

Pothitirat, W., Gritsanapan, W., 2005. Quantitative Analysis of Curcumin, Demethoxycurcumin and Bisdemethoxycurcumin in the Crude Curcuminoid Extract from *Curcuma longa* in Thailand by TLC- Densitometry. Mahidol Univ. J. Pharm. Sci.

Prasad, S., Aggarwal, B., 2011. Turmeric, the Golden Spice. https://doi.org/ 10. 1201/ b10787-14.

Prashina Mol, P., Aparna, R.S., Sheeja, T.E., Deepa, K., 2021. Novel bHLH and WD40 transcription factors from turmeric (*Curcuma longa* L.) as putative regulators of curcumin biosynthesis. J. Plant. Crop. 20–27. https://doi.org/10.25081/ JPC. 2021.V49.I1.7057.

Pretorius, C.J., Tugizimana, F., Steenkamp, P.A., Piater, L.A., Dubery, I.A., 2021. Metabolomics for biomarker discovery: Key signatory metabolic profiles for the identification and discrimination of oat cultivars. Metabolites. https://doi.org/ 10.3390/metabo11030165.

Priyadarsini, K.I., 2014a. The Chemistry of Curcumin: From Extraction to Therapeutic Agent 20091–20112. https://doi.org/10.3390/molecules19122009.

Priyadarsini, K.I., 2014b. The chemistry of curcumin: From extraction to therapeutic agent. Molecules. https://doi.org/10.3390/molecules191220091.

Prosser, G.A., Larrouy Maumus, G., Carvalho, L.P.S., 2014. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. EMBO Rep. https://doi.org/10.15252/embr.201338283.

Purseglove, J. W. 1968. Tropical Crops: Dicotyledons. Volumes 1 and 2. Tropical Crops: Dicotyledons. Volumes 1 and 2.

Fuller, H. R., and Morris, G. E. 2012. Quantitative proteomics using iTRAQ labeling and mass spectrometry. Integrative proteomics, 442, 347-362.https://doi.org/10.5772/31469.

Raes, J., Rohde, A., Christensen, J.H., Van De Peer, Y., Boerjan, W., 2003. Genome-Wide Characterization of the Lignification Toolbox in Arabidopsis. Plant Physiol. https://doi.org/10.1104/pp.103.026484.

Rahi, S., Yadav, D.S., Sood, P., Sharma, K., Sharma, L.K., 2020. Performance of Annually Harvested Improved Turmeric (*Curcuma longa* L.) Cultivars Grown

under Rainfed Conditions. Int. J. Curr. Microbiol. Appl. Sci. https://doi.org/ 10.20546/ ijcmas.2020.906.375.

Rahmani, A., Alsahli, M., Aly, S., Khan, M., Aldebasi, Y., 2018. Role of Curcumin in Disease Prevention and Treatment. Adv. Biomed. Res. https://doi.org/10.4103/abr.ab 147_16.

Rama Rao, M., Rao, D.V.., 1994. Genetic resources of turmeric, advances in horticulture. In. Plant. spice Crop. 9.

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. https://doi.org/10.1016/j.phytochem.2006.06.028.

Rasmussen, S., Dixon, R.A., 1999. Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. Plant Cell. https://doi.org/10.1105/tpc.11.8.1537.

Rattis, B.A.C., Ramos, S.G., Celes, M.R.N., 2021. Curcumin as a Potential Treatment for COVID-19. Front. Pharmacol. https://doi.org/10.3389/fphar.2021.675287.

Ravindran, P.N., Babu, K.N., Sivaraman, K., 2007. Turmeric: The genus *Curcuma*, Turmeric.

Ray, T.A., Cochran, K., Kozlowski, C., Wang, J., Alexander, G., Cady, M.A., Spencer, W.J., Ruzycki, P.A., Clark, B.S., Laeremans, A., He, M.X., Wang, X., Park, E., Hao, Y., Iannaccone, A., Hu, G., Fedrigo, O., Skiba, N.P., Arshavsky, V.Y., Kay, J.N., 2020. Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. Nat. Commun. 11, 1–20. https://doi.org/10.1038/s41467-020-17009-7.

Reddy, N.R.R., Mehta, R.H., Soni, P.H., Makasana, J., Gajbhiye, N.A., Ponnuchamy, M., Kumar, J., 2015. Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties. PLoS One. https://doi.org/10.1371/journal.pone.0129422.

Reichert, A.I., He, X.Z., Dixon, R.A., 2009. Phenylalanine ammonia-lyase (PAL) from tobacco (*Nicotiana tabacum*): Characterization of the four tobacco *PAL* genes and active heterotetrameric enzymes. Biochem. J. https://doi.org/10.1042/BJ20090620.

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. https://doi.org/10.1016/j.gene.2011.09.025.

Rice, P., Longden, L., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. https://doi.org/10.1016/S0168-9525(00)02024-2.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G.L., 2000. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science (80-.). https://doi.org/10.1126/science.290.5499.2105.

Ritter, H., Schulz, G.E., 2004. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. Plant Cell. https://doi.org/10.1105/tpc.104.025288.

Rizvi, S.M.D., Shakil, S., Haneef, M., 2013. A simple click by click protocol to perform docking: Autodock 4.2 made easy for non-bioinformaticians. EXCLI J. https://doi.org/10.17877/DE290R-11534.

Robinson, M.M., Zhang, X., 2011. The World Medicines Situation 2011 Traditional Medicines: Global situation, Issues and Challenges. World Heal. Organ.

Rocha, F.A.C., de Assis, M.R., 2020. Curcumin as a potential treatment for COVID-19. Phyther. Res. https://doi.org/10.1002/ptr.6745.

Rombel, I.T., Sykes, K.F., Rayner, S., Johnston, S.A., 2002. ORF-FINDER: A vector for high-throughput gene identification. Gene. https://doi.org/10.1016/S0378-1119(01)00819-8.

Rösler, J., Krekel, F., Amrhein, N., Schmid, J., 1997. Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. Plant Physiol. https://doi.org/10.1104/pp.113.1.175.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J., 2004. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics. https://doi.org/10.1074/mcp.M400129-MCP200.

Rossmann, M.G., Moras, D., Olsen, K.W., 1974. Chemical and biological evolution of a nucleotide-binding protein. Nature. https://doi.org/10.1038/250194a0.

Röther, D., Poppe, L., Morlock, G., Viergutz, S., Rétey, J., 2002. An active site homology model of phenylalanine ammonia-lyase from *Petroselinum crispum*. Eur. J. Biochem. https://doi.org/10.1046/j.1432-1033.2002.02984.

Roughley, P.J., Whiting, D.A., 1973. Experiments in the biosynthesis of curcumin. J. Chem. Soc. Perkin Trans. 1. https://doi.org/10.1039/p19730002379.

Rushton, P.J., Somssich, I.E., Ringler, P., Shen, Q.J., 2010. WRKY transcription factors. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2010.02.006.

Sablowski, R.W.M., Moyano, E., Culianez-Macia, F.A., Schuch, W., Martin, C., Bevan, M., 1994. A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. EMBO J. https://doi.org/10.1002/j.1460-2075.1994.tb06242.

Sadowski, J., Gasteiger, J., 1993. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. Chem. Rev. https://doi.org/ 10.1021/ cr00023a012.

Sadowski, J., Gasteiger, J., Klebe, G., 1994. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. J. Chem. Inf. Comput. Sci. https://doi.org/10.1021/ci00020a039.

Saga, H., Ogawa, T., Kai, K., Suzuki, H., Ogata, Y., Sakurai, N., Shibata, D., Ohta, D., 2012. Identification and characterization of ANAC042, a transcription factor family gene involved in the regulation of camalexin biosynthesis in Arabidopsis. Mol. Plant. Microbe. Interact. 25, 684–696. https://doi.org/10.1094/MPMI-09-11-0244.

Sajitha, P., Prasath, D., Sasikumar, B., 2014. Phenological variation in two species of *Curcuma*. J. Plant. Crop.

Salim, V., Jones, A.D., DellaPenna, D., 2018. *Camptotheca acuminata* 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism. Plant J. https://doi.org/10.1111/tpj.13936.

Sandeep, I.S., Kuanar, A., Akbar, A., Kar, B., Das, S., Mishra, A., Sial, P., Naik, P.K., Nayak, S., Mohanty, S., 2016. Agroclimatic zone based metabolic profiling of turmeric (*Curcuma Longa* L.) for phytochemical yield optimization. Ind. Crops Prod. https://doi.org/10.1016/j.indcrop.2016.03.007.

Santhi, R., Sheeja, T.E., 2015. Towards computational prediction of microRNA function and activity in turmeric (*Curcuma longa* L.). Indian J. Biotechnol.

Santhi, R., Sheeja, T.E., Krishnamurthy, K.S., 2016. Transcriptome deep sequencing, identification of novel microRNAs and validation under drought stress in turmeric (*Curcuma longa* L.). Plant Biotechnol. Rep. https://doi.org/10.1007/s11816-016-0399-2.

Sasikumar, B., 2005a. Genetic resources of *Curcuma*: diversity, characterization and utilization. Plant Genet. Resour. 3, 230–251. https://doi.org/10.1079/pgr200574.

Sasikumar, B., 2005b. Genetic resources of *Curcuma*: diversity, characterization and utilization. Plant Genet. Resour. https://doi.org/10.1079/pgr200574.

Sasikumar, B., George, J. K., Zachariah, T. J., Ratnambal, M. J., Babu, K. N., and

Ravindran, P. N., 1996. IISR Prabha and IISR Prathibha two new high yielding and high-quality turmeric (*Curcuma longa* L.) varieties. J. Spices and Aromatic crops, 5, 41-48.

Schaart, J.G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A.M.M.L., de Vos, R.C.H., Jonker, H.H., Xu, W., Routaboul, J.-M., Lepiniec, L., Bovy, A.G., 2013. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits. New Phytol. 197, 454–467. https://doi.org/10.1111/nph.12017.

Schaart, J.G., Dubos, C., Romero De La Fuente, I., van Houwelingen, A.M.M.L., de Vos, R.C.H., Jonker, H.H., Xu, W., Routaboul, J.-M., Lepiniec, L., Bovy, A.G., 2013. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits. New Phytol. 197, 454–467. https://doi.org/10.1111/nph.12017.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (80). https://doi.org/10.1126/science.270.5235.467.

Schmidlin, L., Poutaraud, A., Claudel, P., Mestre, P., Prado, E., Santos-Rosa, M., Wiedemann-Merdinoglu, S., Karst, F., Merdinoglu, D., Hugueney, P., 2008. A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. Plant Physiol. https://doi.org/10.1104/pp.108.126003.

Schmittgen, T.D., Livak, K.J., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods.

Schröder, J., 1997. A family of plant-specific polyketide synthases: Facts and predictions. Trends Plant Sci. https://doi.org/10.1016/S1360-1385(97)01104-7.

Schuster, B., Rétey, J., 1995. The mechanism of action of phenylalanine ammonia-lyase: The role of prosthetic dehydroalanine. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.92.18.8433.

Scossa, F., Benina, M., Alseekh, S., Zhang, Y., Fernie, A.R., 2018. The Integration of Metabolomics and Next-Generation Sequencing Data to Elucidate the Pathways of Natural Product Metabolism in Medicinal Plants. Planta Med. https://doi.org/10.1055/a-0630-1899.

Seca, A.M.L., Pinto, D.C.G.A., 2019. Biological Potential and Medical Use of Secondary Metabolites. Medicines. https://doi.org/10.3390/medicines6020066.

Seo, S.W., Bae, G.I.S., Kim, S.G., Yun, S.W., Kim, M.S., Yun, K.J., Park, R.K., Song, H.J., Park, S.J., 2011. Protective effects of *Curcuma longa* against cerulein-

induced acute pancreatitis and pancreatitis-associated lung injury. Int. J. Mol. Med. https://doi.org/10.3892/ijmm.2010.548.

Shakeri, A., Ward, N., Panahi, Y., Sahebkar, A., 2018. Anti-Angiogenic Activity of Curcumin in Cancer Therapy: A Narrative Review. Curr. Vasc. Pharmacol. https://doi.org/10.2174/1570161116666180209113014.

Shanmugasundaram, K.A., Thangaraj, T., Azhakiamanavalan, R.S., Ganga, M., 2001. Evaluation and selection of turmeric (*Curcuma longa* L.) genotypes. J. Spices Aromat. Crop.

Sharifi-Rad, J., Rayess, Y. El, Rizk, A.A., Sadaka, C., Zgheib, R., Zam, W., Sestito, S., Rapposelli, S., Neffe-Skoci ska, K., Zieli ska, D., Salehi, B., Setzer, W.N., Dosoky, N.S., Taheri, Y., El Beyrouthy, M., Martorell, M., Ostrander, E.A., Suleria, H.A.R., Cho, W.C., Maroyi, A., Martins, N., 2020. Turmeric and Its Major Compound Curcumin on Health: Bioactive Effects and Safety Profiles for Food, Pharmaceutical, Biotechnological and Medicinal Applications. Front. Pharmacol. https://doi.org/10.3389/fphar.2020.01021.

Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., Trivedi, P.K., 2016. MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development. Plant Physiol. https://doi.org/10.1104/pp.15.01831.

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 Mol. Biol. Report. https://doi.org/10.1007/s11105-015-0878-6.

Shi, Q., Li, Xi, Du, J., Li, Xingang, 2019. Anthocyanin synthesis and the expression patterns of bHLH transcription factor family during development of the Chinese Jujube fruit (*Ziziphus jujuba* Mill.). Forests. https://doi.org/10.3390/f10040346.

Shoji, T., & Hashimoto, T. 2011. Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by way of the NIC2-locus ERF genes. *Plant and Cell Physiology*, 52(6): 1117-1130.

Sigrist, C.J.A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., Bucher, P., 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief. Bioinform. https://doi.org/10.1093/bib/3.3.265.

Singh, S., Joshi, R.K., Nayak, S., 2013. Identification of elite genotypes of turmeric through agroclimatic zone-based evaluation of important drug yielding traits. Ind. Crops Prod. https://doi.org/10.1016/j.indcrop.2012.07.006.

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. https://doi.org/10.1371/journal.pone.0104005.

Skorobogatko, Y. V., Deuso, J., Adolf-Bergfoyle, J., Nowak, M.G., Gong, Y., Lippa, C.F., Vosseller, K., 2011. Human Alzheimer's disease synaptic O-GlcNAc site mapping and iTRAQ expression proteomics with ion trap mass spectrometry. Amino Acids 40, 765–779. https://doi.org/10.1007/S00726-010-0645-9.

Smith, T.F., Gaitatzes, C., Saxena, K., Neer, E.J., 1999. The WD repeat: A common architecture for diverse functions. Trends Biochem. Sci. https://doi.org/10.1016/S0968-0004(99)01384-5.

Soltani Howyzeh, M., Sadat Noori, S.A., Shariati, V., Amiripour, M., 2018. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. Sci. Rep. https://doi.org/10.1038/s41598-018-31618-9.

Song, J., Wang, Z., 2009. Molecular cloning, expression and characterization of a phenylalanine ammonia-lyase gene (SmPAL1) from *Salvia miltiorrhiza*. Mol. Biol. Rep. https://doi.org/10.1007/s11033-008-9266-8.

Sprenger, H., Erban, A., Seddig, S., Rudack, K., Thalhammer, A., Le, M.Q., Walther, D., Zuther, E., Köhl, K.I., Kopka, J., Hincha, D.K., 2018. Metabolite and transcript markers for the prediction of potato drought tolerance. Plant Biotechnol. J. https://doi.org/10.1111/pbi.12840.

Srinivasan, V., Thankamani, C.K., Dinesh, R., Kandiannan, K., Zachariah, T.J., Leela, N.K., Hamza, S., Shajina, O., Ansha, O., 2016. Nutrient management systems in turmeric: Effects on soil quality, rhizome yield and quality. Ind. Crops Prod. https://doi.org/10.1016/j.indcrop.2016.03.027.

Steinfath, M., Strehmel, N., Peters, R., Schauer, N., Groth, D., Hummel, J., Steup, M., Selbig, J., Kopka, J., Geigenberger, P., van Dongen, J.T., 2010. Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. Plant Biotechnol. J. https://doi.org/10.1111/j.1467-7652.2010.00516.

Stohs, S.J., Chen, O., Ray, S.D., Ji, J., Bucci, L.R., Preuss, H.G., 2020. Highly bioavailable forms of curcumin and promising avenues for curcumin-based research and application: A review. Molecules. https://doi.org/10.3390/molecules25061397.

Stommel, J.R., Lightbourn, G.J., Winkel, B.S., Griesbach, R.J., 2009. Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. J. Am. Soc. Hortic. Sci. https://doi.org/10.21273/jashs.134.2.244.

Sun, H., Yu, J., Zhang, F., Kang, J., Li, M., Wang, Z., Liu, W., Zhang, J., Yang, Q., Long, R., 2020. ITRAQ-based comparative proteomic analysis of differences in the protein profiles of stems and leaves from two alfalfa genotypes. BMC Plant Biol. 20, 1–14. https://doi.org/10.1186/S12870-020-02671-2/Figures/8

Tanvir, E.M., Hossen, M.S., Hossain, M.F., Afroz, R., Gan, S.H., Khalil, M.I., Karim, N., 2017. Antioxidant properties of popular turmeric (*Curcuma longa*) varieties from Bangladesh. J. Food Qual. https://doi.org/10.1155/2017/8471785.

Teng, S., Keurentjes, J., Bentsink, L., Koornneef, M., & Smeekens, S. 2005. Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. *Plant Physiology*, 139(4): 1840-1852.

Thiebaut, F., Rojas, C.A., Almeida, K.L., Grativol, C., Domiciano, G.C., Lamb, C.R.C., De Almeida Engler, J., Hemerly, A.S., Ferreira, P.C.G., 2012. Regulation of miR319 during cold stress in sugarcane. Plant, Cell Environ. https://doi.org/10.1111/j.1365-3040.2011.02430.

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 Res. https://doi.org/10.1093/nar/22.22.4673.

Thwe, A.A., Kim, J.K., Li, X., Bok Kim, Y., Romij Uddin, M., Kim, S.J., Suzuki, T., Park, N. Il, Park, S.U., 2013. Metabolomic Analysis and Phenylpropanoid Biosynthesis in Hairy Root Culture of Tartary Buckwheat Cultivars. PLoS One. https://doi.org/10.1371/journal.pone.0065349.

Tirumalai, V., Swetha, C., Nair, A., Pandit, A., Shivaprasad, P. V., 2019. MiR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. J. Exp. Bot. https://doi.org/10.1093/jxb/erz264.

Todd, A. T., Liu, E., Polvi, S. L., Pammett, R. T., & Page, J. E. 2010. A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. *The Plant Journal*, **62**(4): 589-600.

Tohge, T., Watanabe, M., Hoefgen, R., Fernie, A.R., 2013. Shikimate and phenylalanine biosynthesis in the green lineage. Front. Plant Sci. https://doi.org/ 10.3389/ fpls.2013.00062.

Treumann, A., Thiede, B., 2010. Isobaric protein and peptide quantification: Perspectives and issues. Expert Rev. Proteomics. https://doi.org/10.1586/epr.10.29.

Tsai, C.J., Harding, S.A., Tschaplinski, T.J., Lindroth, R.L., Yuan, Y., 2006. Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytol. https://doi.org/10.1111/j.1469-8137. 2006.01798.

Turnbull, J.J., Nakajima, J., Welford, R.W.D., Yamazaki, M., Saito, K., Schofield, C.J., 2004. Mechanistic Studies on Three 2-Oxoglutarate-dependent Oxygenases of Flavonoid Biosynthesis. J. Biol. Chem. https://doi.org/10.1074/jbc.m309228200

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, 1–13. https://doi.org/10.1186/1471-2164-15-746.

Vandenbroucke, I.I., Vandesompele, J., De Paepe, A., Messiaen, L., 2001. Quantification of splice variants using real-time PCR. Nucleic Acids Res. 29. https:// doi.org/10.1093/nar/29.13. e 68.

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E.F., Hellens, R.P., 2007. Protocol: A highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods. https://doi.org/10.1186/1746-4811-3-12.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science (80-.). https://doi.org/10.1126/science.270.5235.484.

Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D. L., Hanski, I., Marden, J.H., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol. Ecol. https://doi.org/10.1111/j.1365-294X.2008.03666.

Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., Panopoulos, N., 2007. Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. Biotechnol. J. https://doi.org/10.1002/biot.200700084.

Vierstra, R.D., 2003. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci. https://doi.org/10.1016/S1360-1385(03)00014-1.

Vogt, T., 2010. Phenylpropanoid biosynthesis. Mol. Plant. https://doi.org/ 10.1093/ mp/ssp106.

Vom Endt, D., Kijne, J.W., Memelink, J., 2002. Transcription factors controlling plant secondary metabolism: What regulates the regulators? Phytochemistry. https://doi.org/10.1016/S0031-9422(02)00185-1.

Wal, P., Saraswat, N., Pal, R.S., Wal, A., Chaubey, M., 2019. A Detailed Insight of the Anti-inflammatory Effects of Curcumin with the Assessment of Parameters, Sources of ROS and Associated Mechanisms. Open Med. J. https://doi.org/10.2174/1874220301906010064.

Wall, M.E., Wani, M.C., Cook, C.E., Palmer, K.H., McPhail, A.T., Sim, G.A., 1966. Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from *Camptotheca acuminata*. J. Am. Chem. Soc. https://doi.org/10.1021/ja00968a057.

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. Chin. J. Integr. Med. https://doi.org/10.1007/bf02836532.

Wang, J., Wang, X.R., Zhou, Q., Yang, J.M., Guo, H.X., Yang, L.J., Liu, W.Q., 2016. iTRAQ protein profile analysis provides integrated insight into mechanisms of tolerance to TMV in tobacco (*Nicotiana tabacum*). J. Proteomics. https://doi.org/10.1016/j.jprot.2015.11.009.

Wang, T., Guo, S., Jiang, Y., Zou, J., Yu, K., Khattak, A.N., Tian, E., 2021. Comparative transcriptome and iTRAQ-based proteome analysis in mature leaves of Brassica carinata provides insights into the purple leaf color diversity. J. Hortic. Sci. Biotechnol. 96, 444-455. https://doi.org/10.1080/14620316.2020.1863161.

Wang, T., Guo, S., Yu, K., Zou, J., Jiang, Y., Wan, W., Ye, B., Yang, R., Khattak, A.N., Meng, J., Tian, E., 2019. iTRAQ-based comparative proteomic analysis in mature leaves of Ethiopian mustard (*Brassica carinata*) provides insights into the whole-proteomic profiles and the role of anthocyanin biosynthesis in leaf color diversity. https://doi.org/10.21203/RS.2.14299/V1.

Wang, Y., Bryant, S.H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B.A., Thiessen, P.A., He, S., Zhang, J., 2017. PubChem BioAssay: 2017 update. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw1118.

Wang, Y., Chantreau, M., Sibout, R., Hawkins, S., 2013. Plant cell wall lignification and monolignol metabolism. Front. Plant Sci. https://doi.org/10.3389/fpls.2013.00220.

Wang, Y., Hu, W., Wang, M., 1999. HPLC determination of three curcuminoid constituents in rhizoma *Curcuma*. Yaoxue Xuebao.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. https://doi.org/10.1038/nrg2484.

Wang, Z.B., Chen, X., Wang, W., Cheng, K. Di, Kong, J.Q., 2014. Transcriptome-wide identification and characterization of *Ornithogalum saundersiae* phenylalanine ammonia lyase gene family. RSC Adv. 4, 27159–27175. https://doi.org/ 10.1039/C4RA03385J.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P., McPhail, A.T., 1971. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. J. Am. Chem. Soc. 93, 2325–2327. https:// doi. org/10.1021/ja00738a045.

Wanner, L.A., Li, G., Ware, D., Somssich, I.E., Davis, K.R., 1995. The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. Plant Mol. Biol. https://doi.org/10.1007/BF00020187.

Ward, J.A., Ponnala, L., Weber, C.A., 2012. Strategies for transcriptome analysis in nonmodel plants. Am. J. Bot. https://doi.org/10.3732/ajb.1100334.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018.

SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. https://doi.org/10.1093/nar/gky427.

Weitzel, C., Petersen, M., 2010. Enzymes of phenylpropanoid metabolism in the important medicinal plant *Melissa officinalis* L. Planta. https://doi.org/10.1007/s00425-010-1206.

Whetten, R.W., Sederoff, R.R., 1992. Phenylalanine ammonia-lyase from loblolly pine: Purification of the enzyme and isolation of complementary DNA clones. Plant Physiol. https://doi.org/10.1104/pp.98.1.380.

WHO, 2019. World health Organization (WHO) Global report on traditional and complementary medicine 2019, World Health Organization.

Widiez, T., Hartman, T.G., Dudai, N., Yan, Q., Lawton, M., Havkin-Frenkel, D., Belanger, F.C., 2011. Functional characterization of two new members of the caffeoyl CoA O-methyltransferase-like gene family from *Vanilla planifolia* reveals a new class of plastid-localized O-methyltransferases. Plant Mol. Biol. 76, 475–488. https://doi.org/10.1007/s11103-011-9772-2.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., Bähler, J., 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. https://doi.org/10.1038/ nature07002.

Wilson, G.A., Bertrand, N., Patel, Y., Hughes, J.B., Feil, E.J., Field, D., 2005. Orphans as taxonomically restricted and ecologically important genes. Microbiology. https://doi.org/10.1099/mic.0.28146-0.

Wink, M., 2015. Modes of Action of Herbal Medicines and Plant Secondary Metabolites. Medicines. https://doi.org/10.3390/medicines2030251.

Wongthongdee, N., Inprakhon, P., 2013. Stability of turmeric constituents in natural soaps. ScienceAsia. https://doi.org/10.2306/scienceasia1513-1874.2013.39.477.

Wu, Q., Peng, Z., Zhang, Y., Yang, J., 2018. COACH-D: Improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. Nucleic Acids Res. https://doi.org/10.1093/nar/gky439.

Wu, W.W., Wang, G., Baek, S.J., Shen, R.F., 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J. Proteome Res. https://doi.org/10.1021/pr050405o.

Xia, H., Zhu, L., Zhao, C., Li, K., Shang, C., Hou, L., Wang, M., Shi, J., Fan, S., Wang, X., 2020. Comparative transcriptome analysis of anthocyanin synthesis in black and pink peanut. Plant Signal. Behav. https://doi.org/10.1080/ 15592324. 2020.1721044.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.Y., Wei, L., 2011. KOBAS 2.0: a web server for annotation and identification of

enriched pathways and diseases. Nucleic Acids Res. 39, W316. https://doi.org/ 10.1093 /NAR/GKR483.

Xu, F., Deng, G., Cheng, S., Zhang, W., Huang, X., Li, L., Cheng, H., Rong, X., Li, J., 2012. Molecular cloning, characterization and expression of the phenylalanine ammonia-lyase gene from *Juglans regia*. Molecules 17, 7810–7823. https://doi.org/10.3390/molecules17077810.

Xu, J., Liu, J., Chen, J., 2008. Effect of *Ginkgo biloba* extract combined with amlodipine on quality of life in patients with senium hypertension. Guid. Chin. Med. 6, 41–43.

Yamada, Y., Koyama, T., Sato, F., 2011. Basic helix-loop-helix transcription factors and regulation of alkaloid biosynthesis. Plant Signal. Behav. https://doi.org/ 10.4161/psb.6.11.17599.

Yamagishi, M., Shimoyamada, Y., Nakatsuka, T., and Masuda, K. 2010. Two R2R3-MYB genes, homologs of petunia AN2, regulate anthocyanin biosyntheses in flower tepals, tepal spots and leaves of Asiatic hybrid lily. *Plant and Cell Physiology*, 51(3): 463-474.

Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M., Chiang, V.L., Sederoff, R.R., Zhao, X., 2021. Myb-mediated regulation of anthocyanin biosynthesis. Int. J. Mol. Sci. https://doi.org/10.3390/ijms22063103.

Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., Liu, X., Cao, B., Qiu, Z., 2020. Anthocyanin Fruit encodes an R2R3-MYB transcription factor, SlAN2-like, activating the transcription of SlMYBATV to fine-tune anthocyanin content in tomato fruit. New Phytol. https://doi.org/10.1111/nph.16272.

Yandeau-Nelson, M.D., Lauter, N., Zabotina, O.A., 2015. Advances in metabolomic applications in plant genetics and breeding. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. https://doi.org/10.1079/PAVSNNR201510040.

Yang, C.Q., Fang, X., Wu, X.M., Mao, Y.B., Wang, L.J., Chen, X.Y., 2012. Transcriptional Regulation of Plant Secondary Metabolism. J. Integr. Plant Biol. https://doi.org/10.1111/j.1744-7909.2012.01161.

Yang, M., Wang, Q., Liu, Y., Hao, X., Wang, C., Liang, Y., Chen, J., Xiao, Y., Kai, G., 2021. Divergent camptothecin biosynthetic pathway in *Ophiorrhiza pumila*. BMC Biol. https://doi.org/10.1186/s12915-021-01051.

Yang, R., Zeng, Y., Yi, X., Zhao, L., Zhang, Y., 2015. Small RNA deep sequencing reveals the important role of microRNAs in the halophyte *Halostachys caspica*. Plant Biotechnol. J. https://doi.org/10.1111/pbi.12337.

Yavarpour-Bali, H., Pirzadeh, M., Ghasemi-Kasman, M., 2019. Curcumin-loaded nanoparticles: A novel therapeutic strategy in treatment of central nervous system disorders. Int. J. Nanomedicine. https://doi.org/10.2147/IJN.S208332.

Yetgin, S., Olcay, L., Dibar, E., Conter, V., Masera, G., Valsecchi, M.G., Dacou-Voutetakis, C., Henze, G., Loening, L., Schrappe, M., von Stackelberg, A., Zimmermann, M., Attarbaschi, A., Gadner, H., Mann, G., Brandalise, S.R., Carroll, W.L., Devidas, M., Gaynon, P., Hunger, S., Nachman, J., Janka, G., Stary, J., Gelber, R.D., Bierings, M., Kamps, W.A., Pieters, R., Otten, J., Suciu, S., Viana, M.B., Baruchel, A., Ortega, J.J., Magyarosy, E., Perez, C., Steinberg, D., Tsurusawa, M., Zintl, F., Matsuzaki, A., Eden, T.O.B., Lilleyman, J.S., Richards, S., Steinherz, P.G., Kochupillai, V., Sanchez de Toledo, J., Appelbaum, F.R., Campbell, M., Cheng, C., Pei, D., Pui, C.H., Kukure, P., Nakazawa, S., Elphinstone, T., Evans, V., Gettins, L., Hicks, C., Mackinnon, L., Morris, P., Wade, R., Wise, C., 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 5659 children. Br. J. Haematol. https://doi.org/10.1111/j.1365-2141.2010.08148.

Yeung, A.W.K., Horbaczuk, M., Tzvetkov, N.T., Mocan, A., Carradori, S., Maggi, F., Marchewka, J., Sut, S., Dall'Acqua, S., Gan, R.Y., Tancheva, L.P., Polgar, T., Berindan-Neagoe, I., Pirgozliev, V., Šmejkal, K., Atanasov, A.G., 2019. Curcumin: Total-Scale Analysis of the Scientific Literature. Molecules. https://doi.org/10.3390/molecules24071393.

Zhang, A., Han, D., Wang, Y., Mu, H., Zhang, T., Yan, X., Pang, Q., 2018. Transcriptomic and proteomic feature of salt stress-regulated network in Jerusalem artichoke (*Helianthus tuberosus* L.) root based on de novo assembly sequencing analysis. Planta. https://doi.org/10.1007/s00425-017-2818-1.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., Anderson, T.A., 2006. Conservation and divergence of plant microRNA genes. Plant J. https://doi.org/10.1111/j.1365-313X.2006.02697.

Zhang, D.W., Fu, M., Gao, S.H., Liu, J.L., 2013. Curcumin and diabetes: A systematic review. Evidence-based Complement. Altern. Med. https://doi.org/ 10.1155/2013/636053.

Zhang, J., Zhang, S., Li, H., Du, H., Huang, H., Li, Y., Hu, Y., Liu, H., Liu, Y., Yu, G., Huang, Y., 2016. Identification of transcription factors ZmMYB111 and ZmMYB148 involved in phenylpropanoid metabolism. Front. Plant Sci. https://doi.org/10.3389/fpls.2016.00148.

Zhang, L., Chen, C., Xie, F., Hua, Q., Zhang, Z., Zhang, R., Chen, J., Zhao, J., Hu, G., Qin, Y., 2021. A novel wrky transcription factor hmowrky40 associated with betalain biosynthesis in pitaya (*Hylocereus monacanthus*) through regulating hmocyp76ad1. Int. J. Mol. Sci. https://doi.org/10.3390/ijms22042171.

Zhang, M., Yu, Z., Zeng, D., Si, C., Zhao, C., Wang, H., Li, C., He, C., Duan, J., 2021. Transcriptome and metabolome reveal salt-stress responses of leaf tissues from *Dendrobium officinale*. Biomolecules. https://doi.org/10.3390/biom11050736.

Zhang, Shuangyi, Chen, Y., Zhao, L., Li, C., Yu, J., Li, T., Yang, W., Zhang, Shengnan, Su, H., Wang, L., 2020. A novel NAC transcription factor, MdNAC42, regulates anthocyanin accumulation in red-fleshed apple by interacting with MdMYB10. Tree Physiol. https://doi.org/10.1093/treephys/tpaa004.

Zhang, X., Gou, M., Liu, C.J., 2013. Arabidopsis kelch repeat F-Box proteins regulate phenylpropanoid biosynthesis *via* controlling the turnover of phenylalanine ammonia-lyase. Plant Cell. https://doi.org/10.1105/tpc.113.119644.

Zhang, X., Liu, C.J., 2015. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. Mol. Plant. https://doi.org/10.1016/j.molp.2014.11.001.

Zhao, D., Tang, Y., Xia, X., Sun, J., Meng, J., Shang, J., Tao, J., 2019. Integration of Transcriptome, Proteome, and Metabolome Provides Insights into How Calcium Enhances the Mechanical Strength of *Herbaceous peony* Inflorescence Stems. Cells 8, 102. https://doi.org/10.3390/CELLS8020102.

Zhao, J., Sun, C., Shi, F., Ma, S., Zheng, J., Du, X., Zhang, L., 2021. Comparative transcriptome analysis reveals sesquiterpenoid biosynthesis among 1-, 2- and 3-year-old *Atractylodes chinensis*. BMC Plant Biol. https://doi.org/10.1186/s12870-021-03131-1.

Zhao, M., Li, J., Zhu, L., Chang, P., Li, L., Zhang, L., 2019. Identification and characterization of MYB-bHLH-WD40 regulatory complex members controlling anthocyanidin biosynthesis in blueberry fruits development. Genes (Basel). https://doi.org/10.3390/genes10070496.

Zhao, M., Zhong, Q., Tian, M., Han, R., Ren, Y., 2020. Comparative transcriptome analysis reveals differentially expressed genes associated with the development of Jerusalem artichoke tuber (*Helianthus tuberosus* L.). Ind. Crops Prod. 151, 112455. https://doi.org/10.1016/J.INDCROP.2020.112455.

Zhao, N., Wang, G., Norris, A., Chen, X., Chen, F., 2013. Studying Plant Secondary Metabolism in the Age of Genomics. CRC. Crit. Rev. Plant Sci. https://doi.org/10.1080/07352689.2013.789648.

Zhao, Q., Dixon, R.A., 2011. Transcriptional networks for lignin biosynthesis: More complex than we thought? Trends Plant Sci. 16, 227–233. https://doi.org/ 10.1016/j.tplants.2010.12.005.

Zheng, J., Wu, H., Zhu, H., Huang, C., Liu, C., Chang, Y., Kong, Z., Zhou, Z., Wang, G., Lin, Y., Chen, H., 2019. Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. New Phytol. https://doi.org/ 10.1111/nph.15807.

Zhong, R., Ye, Z.H., 2007. Regulation of cell wall biosynthesis. Curr. Opin. Plant Biol. https://doi.org/10.1016/j.pbi.2007.09.001.

Zhou, M., Sun, Z., Wang, C., Zhang, X., Tang, Y., Zhu, X., Shao, J., Wu, Y., 2015. Changing a conserved amino acid in R2R3-MYB transcription repressors results in cytoplasmic accumulation and abolishes their repressive activity in Arabidopsis. Plant J. https://doi.org/10.1111/tpj.13008.

Zhou, S.M., Chen, L.M., Liu, S.Q., Wang, X.F., Sun, X.D., 2015. De novo assembly and annotation of the Chinese chive (*Allium tuberosum* Rottler ex Spr.) transcriptome using the Illumina platform. PLoS One. https://doi.org/10.1371/ journal. pone.0133312.

Zhou, T., Luo, X., Zhang, C., Xu, X., Yu, C., Jiang, Z., Zhang, L., Yuan, H., Zheng, B., Pi, E., Shen, C., 2019. Comparative metabolomic analysis reveals the variations in taxoids and flavonoids among three Taxus species. BMC Plant Biol. https://doi.org/10.1186/s12870-019-2146-7.

Zhou, Y., Gao, F., Liu, R., Feng, J., Li, H., 2012. De novo sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. BMC Genomics 13, 1–13. https://doi.org/10.1186/1471-2164-13-266/FIGURES/7.

Zhu, H.G., Cheng, W.H., Tian, W.G., Li, Y.J., Liu, F., Xue, F., Zhu, Q.H., Sun, Y.Q., Sun, J., 2018. iTRAQ-based comparative proteomic analysis provides insights into somatic embryogenesis in *Gossypium hirsutum* L. Plant Mol. Biol. https://doi.org/10.1007/s11103-017-0681.

Zhu, Q., Xie, X., Lin, H., Sui, S., Shen, R., Yang, Z., Lu, K., Li, M., Liu, Y.G., 2015. Isolation and functional characterization of a phenylalanine ammonia-lyase gene (Ss*PAL1*) from coleus (*Solenostemon scutellarioides* (L.) Codd). Molecules 20, 16833–16851. https://doi.org/10.3390/molecules200916833.

Zimmermann, A., Hahlbrock, K., 1975. Light-induced changes of enzyme activities in parsley cell suspension cultures. Purification and some properties of phenylalanine ammonia-lyase (E.C. 4.3.1.5). Arch. Biochem. Biophys. https://doi.org/10.1016/ 0003- 9861(75) 90364-1.