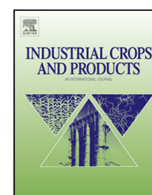
Industrial Crops and Products 97 (2017) 229–241

Contents lists available at ScienceDirect

Industrial Crops and Products

journal homepage: www.elsevier.com/locate/indcrop



Highly conserved sequence of *ClPKS11* encodes a novel polyketide synthase involved in curcumin biosynthesis in turmeric (*Curcuma longa* L.)

Deepa K[a,c], Sheeja TE[a,*], Rosana OB[a], Srinivasan V[b], Krishnamurthy KS[b], Sasikumar B[a]

[a] *Division of Crop Improvement and Biotechnology, ICAR-Indian Institute of Spices Research, Kozhikode, Kerala 673012, India*
[b] *Division of Crop Protection, ICAR-Indian Institute of Spices Research, Kozhikode, Kerala 673012, India*
[c] *University of Calicut, Malappuram, Kerala 673635, India*

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form
30 November 2016
Accepted 5 December 2016

Keywords:
Curcumin
Polyketide synthase
Full length cDNA
Docking
Transcriptome
Gene expression
Open reading frame

ABSTRACT

In order to elucidate a gene regulation model for biosynthesis of the major pharmaceutical compound curcumin in turmeric (*Curcuma longa*), a precise knowledge of sequence diversity and expression patterns of key genes of the pathway is necessary. Polyketide synthases (PKS) being the key enzymes involved in the pathway, attempts were made to mine the major *PKS* from the transcriptome of the *Curcuma* rhizome. Comparative expression of candidate genes *vis a vis* curcumin content across accessions, various developmental stages, environmental conditions and management practices was analyzed. The full length cDNA of a novel *PKS*, showing higher transcript abundance and significant correlation with curcumin content was amplified and bioinformatic analysis was carried out. The present study could mine 63 transcripts of *PKS* from *Curcuma* transcriptome and among them, a novel transcript (*ClPKS11*) showed 69 fold higher expression in a high curcumin variety. The expression of *ClPKS11* correlated with curcumin content under different experimental conditions. It contained an open reading frame of 1176 bp, encoding a polypeptide of 391 amino acids with a predicted molecular mass of 42.9 kDa. CLPKS11 showed maximum identity of 72% with CURS3 (curcumin synthase 3) and exhibited amino acid differences in the substrate binding pocket, cyclization pocket and geometry shapers surrounding the active site. Molecular docking studies indicated a high substrate affinity for CLPKS11. Intrinsic levels of *ClPKS11* may be used as a marker for screening for curcumin, as it shows divergent expressions in high and low curcumin genotypes that are detectable even at the very early developmental stage. The present study also laid the foundation for over expression of *ClPKS11* in turmeric to investigate its physiological role in curcumin biosynthesis.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Curcuma longa L., (turmeric) is an important rhizomatous medicinal spice belonging to the family Zingiberaceae. It is an indispensable ingredient in many Asian dishes that gives yellow colour to the curries and has religious significance. It is also used in traditional medicines since ages to treat a large number of disorders and diseases. Therapeutic properties of turmeric are attributed to the curcuminoid, which is a mixture of curcumin, demethoxycurcumin and bisdemethoxycurcumin. In modern medicine, over 1000 published *in vivo* and *in vitro* studies revealed the importance of curcumin (diferuloylmethane) in treating various diseases ranging from arthritis and inflammation to Alzheimer's disease and cancer. Curcumin has been widely studied for anti-inflammatory (Satoskar et al., 1986), anti-cancer (Kuttan et al., 1985), anti-oxidant (Toda et al., 1985), wound healing (Sidhu et al., 1998) and anti-microbial activities (Negi et al., 1999). Of late curcumin supplements are getting great media coverage and several multinational brands are vying with each other in the global market to promote it as a health supplement.

Curcumin is a phenylpropanoid derivative (Roughley and Whiting, 1973; Kita et al., 2008) and belongs to diarylheptanoid which is a small class of plant secondary metabolites. Schröder (1997) proposed that type III PKS are involved in diarylheptanoid biosynthesis. PKSs generate molecular diversity by utilizing different starter molecules and by controlling the final length of the polyketide. PKS belong to family of condensing enzymes and play a key role in the biosynthesis of variety of secondary metabolites including chalcone, stilbene, phloroglucinols, resorcinols, benzophenones, biphenyls, bibenzyls, chromones, acridones,

* Corresponding author.
E-mail addresses: sheeja@spices.res.in, tesheeja@gmail.com (S. TE).

http://dx.doi.org/10.1016/j.indcrop.2016.12.003
0926-6690/© 2016 Elsevier B.V. All rights reserved.

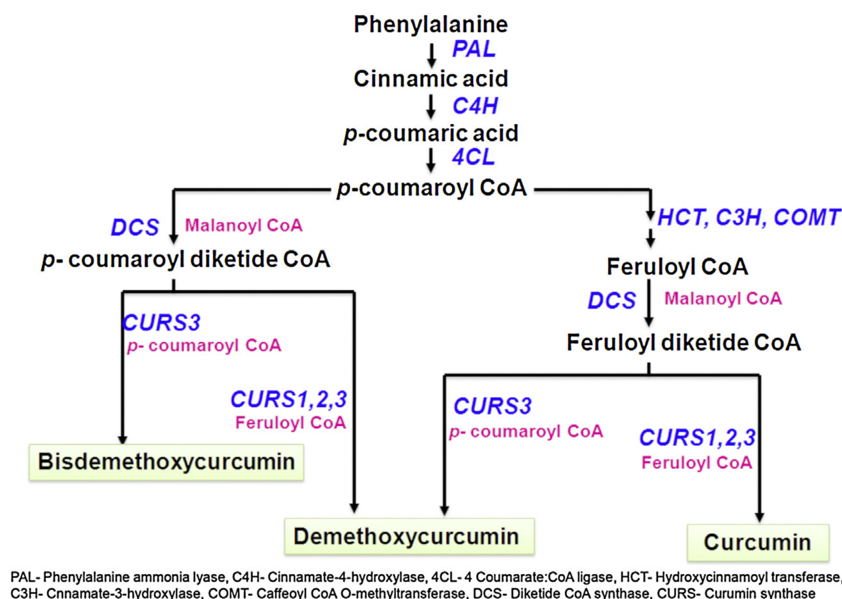


Fig. 1. Proposed pathway for curcumin biosynthesis (Katsuyama et al., 2009a,b).

pyrones etc (Jez et al., 2000). Ramirez-Ahumada et al. (2006) suggested that either a single enzyme or multiple enzymes with similar properties is involved in curcumin biosynthesis. Katsuyama et al. (2009a,b) identified and characterized four type III PKSs from turmeric: one diketide CoA synthase (DCS) and three isoforms of curcumin synthase (CURS). They suggested that DCS synthesizes feruloyldiketide-CoA/*p*-coumaroyldiketide CoA and CURS1-3 then converts diketide-CoA esters into a curcumin scaffold. Further, Katsuyama et al. (2010) exploited *E. coli* carrying an artificially assembled pathway for curcumin biosynthesis. Feeding various analogs of *p*-coumaric acid to *E. coli* harbouring curcumin synthase, acetyl-CoA carboxylase and 4-coumarate CoA ligase led to the production of 17 unnatural curcumin. Resmi and Soniya (2012) suggested that a type III PKS, ClPKS10 may have a possible role in curcumin scaffold biosynthesis. Koo et al. (2013) suggested that the large array of PKS-derived compounds may be the result of multiple PKS-like enzymes catalyzing slightly different reactions, each with different substrate specificities and product outcomes. The involvement of β-ketoacyl-CoA synthase-like subclass, specific reductases, hydroxylases and MYB transcription factors in curcumin biosynthesis has been suggested by Ramirez-Ahumada et al. (2006) and Koo et al. (2013). Thus curcumin biosynthetic pathway can be divided into upstream and downstream sections. The upstream section involves the general phenylpropanoid pathway genes leading to the formation of aromatic CoA esters, *p*-coumaroyl CoA and feruloyl CoA namely *PAL*, *C4H*, *4CL*, *HCT*, *C3H* and *COMT* and the downstream genes include *DCS*, *CURS1*, *CURS2* and *CURS3* (Fig. 1). It has been reported that most of the phenylpropanoid genes exist as multigene families (Nair et al., 2002) and divergent isoforms could play distinct roles in secondary metabolisms (Lu et al., 2006). In *Arabidopsis*, four isoforms of *PAL* and *4CL* have been identified in which *PAL1*, *PAL2* and *4CL3* are known to be involved in flavonoid biosynthesis while *4CL1* and *4CL2* participates in lignin biosynthesis (Li et al., 2015; Huang et al., 2010). Thus specific isoform of phenylpropanoid genes involved in curcumin biosynthesis has to be identified to correlate gene expression and metabolite concentration. Thus at present a detailed research on biosynthetic pathway involving relevant enzymes, rate-limiting step and metabolic regulation is necessary.

Turmeric being a non model crop, for a better understanding of the molecular level information, it was necessary to exploit the high throughput sequencing platforms and bioinformatic tools. The previous study on transcriptome analysis of *C. longa* (high curcumin) and its related species, *C. aromatica* L. (very low curcumin) revealed that all the candidate genes of curcumin biosynthesis were present in both the species. Even though no significant variation was observed in the expression level of candidate genes, many transcripts including *PKS* and transcription factors showing differential expression were identified (Sheeja et al., 2015). The variations in curcumin content in turmeric accessions provide interesting material for the present study. The present study tried to focus on the major *PKS*, being the most important class of key enzymes of the pathway. Moreover, it has been reported that curcumin accumulation often exhibit spatio-temporal and environmental variations. Therefore differential expression analysis between genotypes varying in curcumin content will give a better understanding about genes involved in this pathway. Moreover, the involvement of the key genes in biosynthesis of curcumin may be confirmed through correlation of the gene expression levels with the metabolite levels as done in similar studies (Wei et al., 2011). Hence in continuum with the earlier studies (Sheeja et al., 2015), the expression profile of major *PKS* under different experimental conditions was evaluated and examined their correlation with curcumin content to verify their involvement in curcumin synthesis. The present study also describes the cloning and characterization of a major novel *PKS* (*ClPKS11*) from *C. longa*. These results will contribute to further study on the role of *PKS* genes in curcumin biosynthesis and also provide more insights into pathway modelling and regulation.

2. Materials and methods

2.1. Plant material

To construct full length cDNA library, leaf, pseudostem, rhizome and root of a high curcumin turmeric genotype *viz.*, Mega turmeric maintained in ICAR- Indian Institute of Spices Research (IISR) experimental farm, Peruvannamuzhi were used. Since the curcumin content of IISR Prathibha was reported to be higher in Kozhikode (4.6%) than Coimbatore (2.9%) (Anandaraj et al., 2014), these two locations were chosen to study the influence of environment on *PKS* genes. Two turmeric genotypes with contrasting curcumin content viz., IISR Prtahibha, a released variety of IISR

with high curcumin content and accession 449, a germplasm entry with low curcumin content were planted in the field of ICAR-IISR (11.2994°N, 75.8407°N), Kozhikode, Kerala under rain-fed condition and at Tamil Nadu Agricultural University (9.1913°N, 77.8803°E) Coimbatore, Tamil Nadu under irrigated condition. The fertilizers were applied as per the farm guidelines. The same set of plants were also planted in grow bags filled with potting mixture devoid of fertilizers to limit the curcumin biosynthesis and was maintained in green house at IISR, Kozhikode. Leaves and rhizomes were collected during 60, 120 and 180 days after planting (DAP).

For studies on effect of light regimes, IISR Prathibha was grown under different shade nets (red, green and white) where the open condition served as the control. For studies on the effect of management practices, three treatments viz., (1) Organic 100% (30 t FYM + 2 t Neem cake + 1 t Ash + 4 t) Vermi compost per ha, Biofertilizer – *Azospirillum* and *Pseudomonas sp*. as seed treatment and spray of Bordeaux mixture and neem oil for disease and pest control, (2) Inorganic 100% (Recommended dose of fertilizer NPK at 75, 50, 50 kg/ha with recommended chemical methods of pest and disease control) and (3) Integrated management (20 t FYM + half the recommended N, full P and K, P Solubilising bacteria and spray with Dithane M45 and Quinalphos) were employed. The rhizome samples were collected at 120 dap.

To analyze the sequence diversity of *ClPKS11*, the rhizomes of three high curcumin genotypes (IISR Alleppey supreme, Mega turmeric, IISR Prathibha), three low curcumin genotypes (Acc.19, Acc. 200 and Acc. 449) and *C. aromatica* were collected from IISR experimental farm, Peruvannamuzhi.

2.2. Screening of Curcuma *transcriptome for type III* PKS

Homologous transcripts to type III *PKS* were searched for within the text of annotated genes in *C. longa* and *C. aromatica* transcriptome data (Sheeja et al., 2015; Santhi et al., 2016). Gene ontology terms for each unigene were retrieved using the default settings of Blast2GO software (Conesa and Gotz, 2008).

2.3. Total RNA isolation

Total RNA was isolated as described in Deepa et al., 2014 from a pool of three biological replicates for each tissue and DNA digestion was done with DNase I (Qiagen). The quality of RNA was checked in 1.2% agarose and quantified using Biophotometer plus (Eppendorf, Germany). The samples used had two discrete bands of 28S and 18S rRNA on agarose gel and $A_{260}/_{280}$ ratio between 1.8 and 2.0.

2.4. Quantitative real-time PCR

First strand cDNA was synthesized from 250 ng of total RNA 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 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 0.4 μl of cDNA in a final volume of 20 μl. PCR amplification was performed under the following conditions: 95 °C for 5 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30s. A melt curve program of 65–99 °C was included to check the specificity of PCR products. Three technical replicates of each reaction were performed and 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 to check non-specific amplification respectively. The primers were designed based on previous transcriptome data (Sheeja et al., 2015) using Primer Quest (www.idtdna.com/primer quest/home/index) and listed in Table 1. The primers for the curcumin biosynthetic pathway genes were designed based on the contigs which were up-regulated in *C. longa*. The stability of six candidate reference genes were evaluated using a web based analysis tool, RefFinder (Xie et al., 2012) which integrates four computational programs namely GeNorm, Normfinder, BestKeeper and comparative delta Ct method to compare and rank the candidate reference genes. Standard curves were generated for each gene with serial dilutions of pooled cDNAs (10^{-1} to 10^{-6}) to calculate the PCR efficiency. The relative gene expression was calculated using the $2^{-D\Delta Ct}$ method (Livak and Schmittgen 2001) and transformed to $\log_2$ scale.

2.5. Cloning of full length cDNA of ClPKS11

The full length cDNA of *ClPKS11* was amplified using inverse PCR approach as described by Kuniyoshi et al., 2006. Briefly, first strand cDNA was synthesized from 1 μg of pooled total RNA (250 ng each from four turmeric tissues *viz.,* leaf, pseudostem, rhizome and root) using SMARTer PCR cDNA synthesis kit (Clontech) according to manufacturer's instruction. Double-stranded cDNA was prepared using PrimeSTAR HS DNA polymerase (Takara) and circularized with T4 DNA ligase (Promega). Inverse PCR was carried out with circularized double stranded cDNA using Ex-Taq DNA polymerase and outward designed gene specific primers IP1 and IP2. A second round of PCR was carried out with nested primers IP3 and IP4 (Table 1) with same PCR conditions to increase the specificity of PCR products. These primers were designed from a transcript (326 bp) which showed 69 fold higher expression in *C. longa* compared with *C. aromatica* on comparative transcriptome analysis. The PCR products were purified, ligated in pGEM-T vector and transformed in *E. coli* JM109 cells. The positive recombinants were sequenced bi-directionally using M13 primers. The sequence analysis was carried out in Bioedit software (Hall, 1999) with the known sequence of *ClPKS11*. The full length coding sequence was further confirmed by end to end PCR with *ClPKS11* full length primers. The same primer pair was used to generate the genomic DNA sequence of *ClPKS11*.

2.6. Sequence analysis of ClPKS11

Deduced amino acid sequences of *ClPKS11* was obtained by translating the cDNA sequences using the EMBOSS program TRANSEQ (Rice et al., 2000). Comparative sequence analysis of *ClPKS11* was performed online using blastp (Altschul et al., 1990). The open reading frame (ORF) was predicted by ORF Finder (Rombel et al., 2002). The physical and chemical parameters of the protein sequences were analysed using Protparam tool (Gasteiger et al., 2005). Subcellular localisation analysis and the presence of signal peptide and transmembrane regions were analysed using SLP-local (Matsuda et al., 2005), SignalP 4.1 server (Petersen et al., 2011) and TMPred server (Hofmann and Stoffel, 1993) respectively. Multiple sequence alignment of protein sequences with other PKSs was conducted with ClustalW (Thompson et al., 1994) using default parameters. The amino acid residues of substrate binding pocket, cyclization pocket and geometry shapers were searched using the SBSPKS web server (Anand et al., 2010).

2.7. Phylogenetic analysis

Phylogenetic analysis for sequences was performed using different methods such as Bayesian analysis, maximum parsimony, and maximum likelihood and the consense tree was taken. Bayesian analysis was performed in MrBayes version 3.1 (Ronquist and Huelsenbeck, 2003) with two searches run simultaneously for atleast two million generations. Flat-Dirichlet priors were used for the gamma shape parameter and the proportion of invariable sites. Three heated chains (temperature 0.2) and one cold chain were

Table 1
Primers used for quantitative real time PCR and full length gene amplification.

Quantitative real time PCR primers to amplify candidate reference genes				
	Forward primer (5′-3′)	Reverse primer (5′-3′)	Regression coefficient (r^2)	PCR efficiency (%)
EF1α	GCTGACTGTGCTGTTCTCATTAT	CTCGTGTCTGTCCATCCTTTGAA	0.988	101
UBIQUITIN	GCACTCTCGCTGACTACAAC	GGCTTGGTGTAGGTCTTCTTC	0.996	116
ACTIN	CAACAGCAGAACGGGAAATTG	CATAATCAAGGGCGACATATGC	0.992	101
18SrRNA	CCTTCCTCTAAATGATAAGGTTCAATG	GATTGAATGGTCCGGTGAAGTGTT	0.970	99
GAPDH	AACTGTAGCCCCACTCATTG	GCATCTTAGGGTATGTGGAGG	0.998	109
TUBULIN	GGCAGAGATCAGATGGTTCAG	TGGACAATGAAGCACTCTACG	0.989	92
Quantitative real time PCR primers to amplify candidate *PKS* genes				
DCS	GTGCTGTTCATCCTGGACGAG	CAACAGCACGCCCCAGTCGA		
CURS1	TCAGCTCATCCATCACGAAGTACAC	CATCATTGACGCCATCGAAGC		
CURS2	TGTTGCCGAACTCGGAGAAGAC	TCGGGATCAGGACTGGAACAAC		
CURS3	CCCATTCCTTGATCCCTTTTCC	TGGAGCCCTCCTTCGACGACC		
ClPKS11	TGTCGGAGATCACCCACTTG	CGGAGAAGGGAGACCGAGA		
Inverse PCR primers				
IP1: TCGTCGAAGGAAGTGTCCAT		IP2: CGATATGCCTGGAGCTGATT		
IP3: GCCAAACTTCTCGGTCTCTC		IP4: GTCCATGTACGAGCACATCC		
Full length primers				
Fp: CAGCAACCAGTTTCGCTTTC		Rp: CGCCGGAATAAGCTACAGAG		

used in each search; the parameter was then fixed for a bootstrap analysis with 10000 replicates. Maximum likelihood analysis was performed using GARLI version 2.0 (Zwickl, 2006) with two replicates used to estimate model parameters; these parameters were then fixed for a bootstrap analysis with 10000 replicates. Maximum parsimony analysis was done with DNApars of Phylip package. The majority-rule consensus of the bootstrap replicates was calculated in consense and seqboot in the Phylip package (Felsenstein, 1993). When consense is run the majority rule consensus tree will result, showing the outcome of the analysis and allows bootstrapping on different methods in the package. *E. coli* 3-oxoacyl-ACP synthase (GenBank: WP_052924124.1) has been designated as out-group for the analysis.

2.8. Molecular modelling

Since experimentally and computationally solved structures were not available for ClPKS11, three-dimensional (3D) protein structure was constructed using Modeller 9.10 package (Eswar et al., 2006). Calculation of cavity volumes was performed with the CASTP program (Dundas et al., 2006) and the 3D molecular structures and active-sites of proteins were visualized with UCSF Chimera (Huang et al., 1996). Template structure of curcumin synthase (PDB ID: 3OV2) was downloaded from RCSB PDB (Bernstein et al., 1977). The side chains and hydrogen atoms added for refining the structure and the stability of homology model has been validated by checking the geometry using PROCHECK (Laskowski et al., 1993). The model was validated using Ramachandran plot. Ramachandran plot was identified by Procheck program of Structural Analysis and Verification Server (Laskowski et al., 1993).

2.9. Molecular docking

Molecular docking study was carried out using Molegro Virtual Docker (Thomsen and Christensen, 2006) with ClPKS11 and CURS as templates. The canonical smiles notations of substrates (*p*-coumaroyl CoA, feruloyl CoA, *p*-coumaroyldiketide CoA and feruloyldiketide CoA) were collected from PubChem (Kim et al., 2015). The 3D structures of compounds were developed by 3D Structure Generator CORINA (Sadowski et al., 2003) using canonical smiles of the compound. Docking was carried out separately for ClPKS11 and CURS using single large volume active-site.

2.10. Estimation of curcumin content

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.

2.11. Statistical analysis

All statistical parameters were analyzed using SPSS 20.0 software. The relationship between gene expression and curcumin accumulation was performed using Pearson's correlation coefficient and subjected to One-way analysis of variance (ANOVA) at the 0.01 significance level.

3. Results and discussion

3.1. Identification of PKS candidates in Curcuma transcriptome

The transcriptome data was highly instrumental in identifying the key *PKS* gene transcripts involved in curcumin biosynthesis. A total of 34 and 29 *PKS* unigenes which includes curcumin synthase, diketide CoA synthase and chalcone synthase from the transcriptomes of *C. longa* and *C. aromatica* respectively were identified. The presence of several *PKS* unigenes may be responsible for the production of large array of diarylheptanoids in turmeric as suggested by Xie et al., 2009. Twelve unigenes were found to be up-regulated and 10 unigenes were found to be down-regulated in *C. longa*. These up-regulated *PKS* unigenes may have a direct relation with curcumin biosynthesis. Four unigenes showed >75% identity to type III *PKS*2 of *Musa acuminata* Colla. with 64–69 fold higher expression in *C. longa*. Similarly, three unigenes showed 94–96% identity to chalcone synthase of *Kaempferia elegans* Wall. with 30–37 fold higher expression in *C. longa*. This suggests that these contigs might be members of multigene families or part of same gene without overlapping sequences to form single large contigs (Upadhyay et al., 2014). The remaining *PKS* unigenes were found to be expressed in both the transcriptomes without much variation in their expression level. Functional annotation of these twelve up-regulated unigenes by gene ontology analysis annotated two unigenes as curcumin synthase (691222 and 652976) and ten unigenes as chalcone synthase (Table 2).

Table 2
Gene ontology analysis of unigenes up-regulated in *C. longa*.

Contig ID	Sequence length	FPKM value in C. longa	*FPKM value in C. aromatica*	Fold change	Sequence description	Blast similarity mean	Blast top hit description	GO names
691222	326	695.55	10.04	69.25	curcumin synthase	88.8	gi|293632012|gb|ADE59486.1|	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process
675683	350	755.69	11.09	68.17	chalcone synthase	93.15	gi|333609077|gb|AEF58784.1|	naringenin-chalcone synthase activity, biosynthetic process
698710	470	621.33	9.28	66.96	chalcone synthase	92.55	gi|695066950|ref|XP_009380378.1|	naringenin-chalcone synthase activity, flavonoid biosynthetic process
723003	284	439.81	6.68	65.85	chalcone synthase	86.85	gi|695066948|ref|XP_009380377.1|	transferase activity, transferring acyl groups other than amino-acyl groups, protein homodimerization activity, flavonoid biosynthetic process
674483	182	755.80	11.82	63.96	chalcone synthase	79.75	gi|20540|emb|CAA32738.1|	naringenin-chalcone synthase activity, flavonoid biosynthetic process
562074	306	137.70	3.74	36.82	chalcone synthase	89	gi|333609095|gb|AEF58793.1|	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process
599314	190	62.54	1.78	35.05	chalcone synthase	87.65	gi|333609095|gb|AEF58793.1|	naringenin-chalcone synthase activity, biosynthetic process
683784	244	133.62	4.07	32.84	curcumin synthase partial	87.35	gi|723782502|gb|AIY23488.1|	transferase activity, transferring acyl groups other than amino-acyl groups, biosynthetic process
535759	962	68.67	2.30	29.80	chalcone synthase	84.85	gi|293632012|gb|ADE59486.1|	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process
608568	216	65.74	3.13	20.98	chalcone synthase	90.95	gi|333609095|gb|AEF58793.1|	naringenin-chalcone synthase activity, protein homodimerization activity, flavonoid biosynthetic process
648985	187	515.31	24.75	20.82	putatative chalcone synthase	92.15	gi|57471148|gb|AAW50921.1|	naringenin-chalcone synthase activity, biosynthetic process
652976	482	58.80	3.44	17.10	chalcone synthase	92.65	gi|695066950|ref|XP_009380378.1|	naringenin-chalcone synthase activity, flavonoid biosynthetic process

Table 3
Stability of reference gene expression in developing turmeric tissues.

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

3.2. Identification of stable reference genes in turmeric

All the six reference genes were expressed in all the tissues of turmeric. Gene specific amplification was confirmed by visualizing a specific amplicon in 2% agarose gel electrophoresis, a single peak in dissociation curve analysis and by sequencing. The PCR efficiency for the reference genes ranged between 92% to 116% and correlation coefficient (r^2) value from 0.970 to 0.996 (Table 1). The stability of reference genes when analyzed using Reffinder, the overall order from the most stable to the least stable reference genes was: *EF1α* <*UBIQUITIN* <*18SrRNA* <*GAPDH* <*ACTIN* <*TUBULIN*. All the computational programs, except BestKeeper identified *EF1α* as the most stable gene in both developmental stages and tissues (Table 3). Except *TUBULIN*, all the other genes had Genorm M value below 1.5, indicating that the expressions of all the genes except tubulin are stable across the samples (Fig. 2). Hence the most stable genes, *EF1α* and *UBIQUITIN* were selected as reference genes for normalizing the expression of target gene. Further stable expression of *EF1α* and *UBIQUITIN* was observed under different light regimes and management practices with Genorm M value below 1.5.

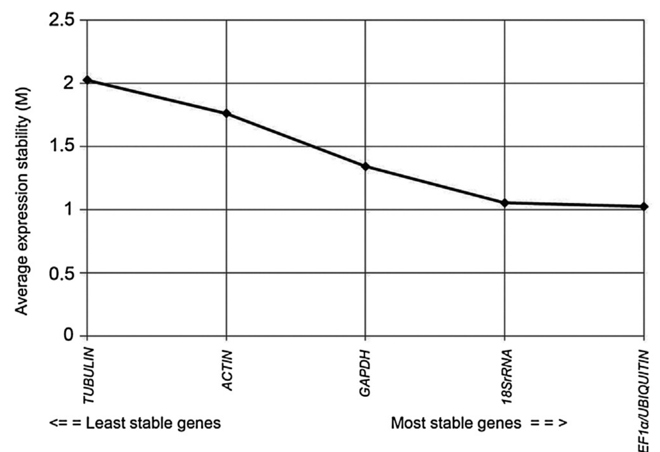


Fig. 2. Average expression stability values (M) of the candidate reference genes.

3.3. *Expression pattern of* PKS *vs curcumin content*

The expression pattern of four reported type III *PKS* in curcumin biosynthesis (*DCS, CURS1, CURS2, CURS3*) and a novel type III *PKS* (*ClPKS11*) were analyzed in the present study. The expression pattern of these genes in leaves and rhizomes of Prathibha planted in field at Kozhikode was studied in three developmental stages (60, 120 and 180 dap) (Fig. 3a and 3b). Compared with leaves, the expression of *CURS1* and *ClPKS11* were up-regulated in rhizomes from 60 to 120 dap and then decreased at 180 dap. The expression of *DCS, CURS2* and *CURS3* showed maximum expression at 120 dap compared with 60 dap and 180 dap. The curcumin content of IISR Prathibha was found to be maximum in rhizomes when harvested at 120 dap (6.52%) compared with rhizomes harvested at 60 dap (0.79%) and 180 dap (5.73%). The result is consistent with the previous biochemical study (Neema, 2005). Thus all the type III *PKS* showed higher expressions at 120 dap in rhizome when curcumin content was maximum (6.52%). 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 were correlated to metabolite accumulation.

Four *PKS* genes namely *DCS, CURS1, CURS3* and *ClPKS11* showed higher expression in IISR Prathibha compared with Acc. 449 (Fig. 3c) when planted in three different environmental conditions. From the same study, it was observed that IISR Prathibha grown in Kozhikode under field conditions showed the highest *PKS* expression (Fig. 3d). The curcumin content of IISR Prathibha was higher under field conditions at Kozhikode (6.52%) than Coimbatore (4.53%) and was lowest under nutrient limiting condition in green house at Kozhikode (1.54%). Similar trend was observed in Acc. 449, which showed highest curcumin content under field conditions at Kozhikode (2.11%) compared with Coimbatore (0.56%) and negligibly low under nutrient limiting conditions in green house at Kozhikode (0.23%). It is thus evident that curcumin is subject to environmental variations and conditions of growth and both high and low curcumin genotypes showed a consistent behaviour. It has been reported that curcumin content varies with varieties (Sasikumar 2005; Sajitha et al., 2014) and subject to agro-climatic variations (Anandaraj et al., 2014; Singh et al., 2013). Our results suggest that the agro-climatic condition in Kozhikode favoured the production of curcumin in IISR Prathibha, when grown under rainfed condition. 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).

The effect of chemical management system on IISR Prathibha decreased the expression of *CURS1-3* and *ClPKS11* compared with organic and integrated management systems (Fig. 3e). Integrated management practice yielded turmeric with high curcumin content (6.15%) compared with those grown under organic (5.71%) and chemical management system (4.79%), highlighting the eminence 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 influence of red, white and green shade nets in turmeric was also assayed in IISR Prathibha where the open condition served as the control. In this study, the expression of all the five *PKS* was found to be reduced in red and white shade net conditions and on par with green suggesting the regulatory role of light in the expression of *PKS* genes and thereby curcumin levels (Fig. 3f). The curcumin content was maximum in control (open- 5.22%) compared with that under green (4.93%), white (4.84%) and red (4.45%) shade nets. The accumulation of anthocyanin was also found to be suppressed by shading (Jeong et al., 2004).

3.4. *Correlation between gene expression and curcumin content*

Among the five *PKS* genes analyzed, *CURS1, CURS2* and *ClPKS11* showed positive correlation with curcumin content when IISR Prathibha was analyzed at 120 dap from different environmental conditions, light regimes and management practices (Table 4). The curcumin content and the expression of *PKS* genes were maximum at 120 dap and was reduced at 180 dap. This might be attributed to the reason that curcumin levels are more or less stable after 180 dap (Neema, 2005). Similar results were also reported in *Matthiola incana* L. where there was a reduction in the activity of biosynthetic enzymes of anthocyanain pathway in the later stages of development; however the anthocyanin content remains constant over time (Dangelmayr et al., 1983). The expression of *CURS2* was down-regulated (0.7 fold) in IISR Prathibha (curcumin content-1.54%) than Acc. 449 (curcumin content-0.23%) under nutrient limiting conditions. So to conclude the expression of *CURS1* and *ClPKS11* correlated with curcumin content in all the experiments in the present study, while *CURS2* did not. Intrinsic levels of *ClPKS11* may be used as a marker for screening for curcumin. A previous study (Sheeja et al., 2015) could identify 5488 putative SSRs from *C. longa* transcriptome and 69 SNPs associating with curcumin content by ddRAD sequencing (unpublished). In future, the association of these markers with the differentially expressed curcumin biosynthetic genes may aid in screening of the elite turmeric genotypes at juvenile stages itself.

3.5. *Amplification of full length* ClPKS11 *cDNA and sequence analysis*

Inverse PCR using double stranded circularized cDNA amplified an intense ~1.6 kbp fragment. The amplicon was sequenced and the resulting sequence was aligned and assembled with the short sequence of *ClPKS11* to obtain the full length cDNA. Finally, an end to end PCR including both start and stop codon of *ClPKS11* was amplified from both cDNA (GenBank accession no. KX017475) and genomic DNA. Sequence analysis indicated that *ClPKS11* ORF is 1176 bp long that corresponded to a deduced protein sequence of 391 amino acid residues with a predicted molecular mass of 42.9 kDa and a pI of 6.11. The ORF region was bordered by 81 bp 5′-untranslated region (UTR) and 262 bp 3′-UTR. A putative polyadenylation signal (AATAA) was identified 236 bp downstream from the stop codon (TAG) in the 3′-UTR. Genomic DNA sequence of *ClPKS11* indicated the presence of an intron of 91 bp (Fig. 4). When the full length cDNA of *ClPKS11* was aligned to rhizome specific transcriptome of *C. longa* and *C. aromatica*, six unigenes showed ≥ 96% identity, suggesting they may be the truncated products or isoforms of *ClPKS11*.

Mallika et al. (2011) reported that plant type III PKSs are localised in cytoplasmic matrix exclusive of transmembrane peptide. ClPKS11 was also found to be localized in cytoplasmic matrix exclusive of any signal peptide. The deduced amino acid sequence of ClPKS11 shared 80% identity with CURS2 like polypeptide of *Musa accuminata*, 72% identity with CURS3, 70% identity with CURS1, 69% identity with CURS2 and 61% identity with DCS of *C. longa*. Thus the comparative sequence analysis suggested that the ClPKS11 is a novel type III PKS.

Multiple sequence alignment of ClPKS11 with other PKSs revealed that the catalytic triad residues (Cys164, His303 and Asn336), identical six residue loops (Thr132-Met137), *cis*-peptide bond between Met137 and Pro138 and gatekeepers (Phe125 and Phe265) to block the lower portion of the opening between the CoA-binding tunnel and the active-site cavity were conserved in the

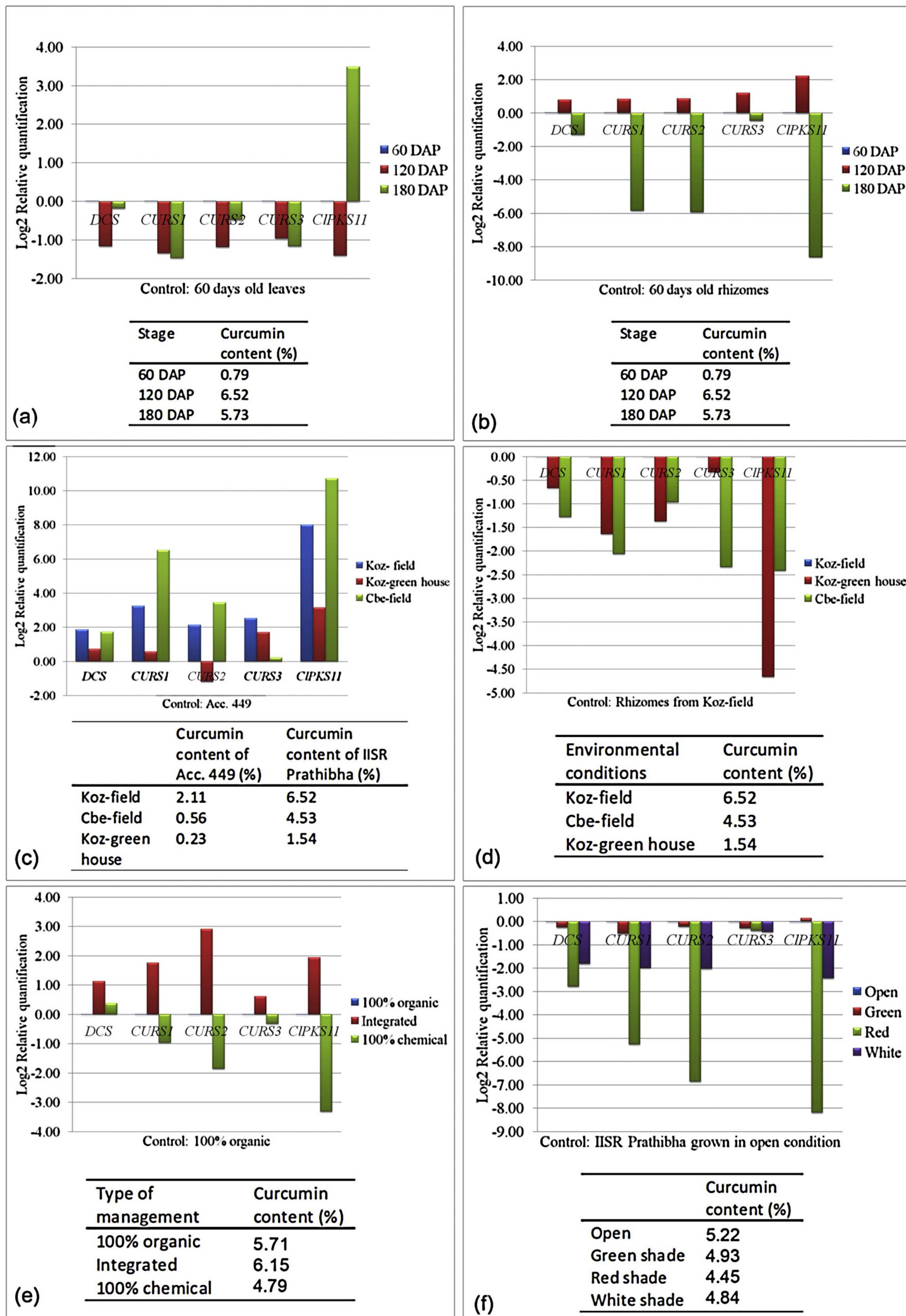


Fig. 3. Relative expression patterns of *PKS* genes in (a) leaves and (b) rhizomes of IISR Prathibha at three developmental stages (c) IISR Prathibha compared with Acc. 449 from three environmental conditions, (d) IISR Prathibha from three environmental conditions (e) different management practices and (f) shade net experiments.

Table 4
Pearson correlation coefficient of *PKS* genes with curcumin content in IISR Prathibha planted in three environmental conditions, different shade net and management practices.

	Pearson correlation coefficient		
	Environmental conditions	Management practices	Shade net experiment
DCS	−0.436	0.829	0.979
CURS1	0.791	0.995	0.986
CURS2	0.819	0.990	0.968
CURS3	−0.355	0.996	0.766
ClPKS11	0.997	0.913	0.964

```
acatggggaaggcatcccattcccatatccctccgagcatcgacgcttgctcgtgcagcaaccagtttcgctttcccctccATGGCCTC
CAACTACGTCGATGCATTCCCCAAGCCCCAAAGGGCTCAAGGCCCAGCCACCGTCATGGCCATCGGCAC
CGCCAACCCTCCCAACCTCTACGAACAGAGCGCTTACCCAGACTTCTATTTCCGGGTCACCGGTGCCGAC
CACAAGCCGGAGCTCAAGCAGAAGTTCCGCCGTCTCTGTaatacaattcatcccgtcgtcgtcgtcgtcgcatactgat
gattcttaatttgttgttacgctaatcgtgttgcgtgctatttgtgcaggtGACAGGAGCATGATCAAGAAGCGTTATATGCA
CCTGACGGAGGAGCTGCTGAAGGAGAAACCAGGGATGTGCTCGTACATGGACACTTCCTTCGACGAGC
GGCAGGATGTCGTGGTGGAGGAGGTGCCTCGCCTGGCCAAGGAGGCCGCCGTCAAGGCCATCAAGGA
GTGGGGACGGCCCTTGTCGGAGATCACCCACTTGGTTTTCTGCTCCACCAGCGGCGTCGATATGCCTGG
AGCTGATTACCGCCTCGCCAAACTTCTCGGTCTCTCCTTCTCCGTCAACCGCATCATGCTCTACAACCAGG
CCTGCCACATCGGCGCGCAGACGCTCCGCATCGCCAAGGACATCGCCGAGAATAACCGGAGCGCCCGC
GTCCTCGTCGTCGCCTGCGAGGTCAACACGCTCATCTTCCGCGGTCCCGAAGAGCGCGACTTCCAGAGC
CTCGCGGCCCAGGTCGCGTTCGGCGACGGAGCGGCGGCGCTCGTCGTCGGGGCCGACCCCGTCCAGG
GCGTCGAGAAGCCGATCTTCGAGATCATGGCGGCGTTTCCGTTCACGGTGCCGGAGACCCAGATGGCG
GTCGGCGGGCAGCTGAAGCAGATCGGGCTGACCTTCCATTTCGCGCACCAGCTGCCGGGGCTGATAGC
CAACAACTTGGAGACGTGCCTCGGCGAGGCGTTGAAGCCGCTGGGGATCTCCGACTGGAACGACGTG
TTCTGGGTGGCCCACCCGGGGAACTGGGGCATCATGGACGCCGTCGAGGCCAAGCTGGGCCTGGAAC
AGGGGAAGCTGCAGTCGTCGAGGCACGTCTTCAGCGAGTTCGGGAACATGATGAGCGCCACCGTTCTG
TTCGTGATGGACGACGTGAGGAAGCGGGCGGTGGCGGAGGGCGCGGCGACCACCGGCGACGGCCTG
AAGTGGGGCGTGCTCTGCGCGTTCGGCCCAGGGCTGTCCATCGAGACGCTGGTGCTTCGCAGCGTCCC
TCTGTAGcttattccggcgccttaccgcccaacatgcttgcataaatattaaaagctgtttaatgccaaaataataatcgattaaatta
attacagtatatcagatcagatgtatacatacgttcgagctgagttgcctgtcacctcccctctgctccctttcgtcatcttaatttcttgatc
acgtcctgttcttgctggtcagcatgttattcaaaagattgataacaattaatgttattaataaaaaaaaaaacaaaaaaaaaaagt
```

Fig. 4. Coding, untranslated and intron sequences of *ClPKS11*. UTR, intron and polyadenylation sequences are depicted in red, blue and yellow boxes respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Amino acid residue changes in the cyclization pocket, substrate binding pocket and geometry shapers.

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

ClPKS11 (Fig. 5a). However, amino acid differences were observed in substrate binding pocket, cyclization pocket, and geometry shapers surrounding the active site (Table 5). The data shows that DCS, CURS1, CURS2, CURS3 and ClPKS10 have similar residues lining the substrate binding pocket and CURS1, CURS2, CURS3 and ClPKS10 share similar cyclization pocket. DCS is having single amino acid change in substrate binding pocket compared with isoforms of curcumin synthase. Notably, in ClPKS11, residues lining both substrate binding pocket and cyclization pocket were different from other type III PKSs reported from *C. longa*. The residues Thr132, Ser133, Thr194, Thr197, Gly256, Phe265 and Ser338 (numbering in *Medicago sativa* L. CHS2) are thought to be critical for controlling the substrate and product specificity of the enzyme reactions (Abe and Morita, 2010). In ClPKS11, Thr132, Ser133, Gly256 and Phe265 were conserved, however Thr194, Thr197 and Ser338 are specifically replaced by Asp, Iso and Met respectively in the corresponding positions of ClPKS11. Mutational studies reported that Thr197 and Ser338 play a role in the polyketide chain elongation reactions (Jez et al., 2000; Abe et al., 2006). When Ser338 was replaced by valine, in a mutant of *Rheum palmatum* L., benzylacetone synthase exhibited a 2-fold increase in bezalacetone-forming activity (Abe et al., 2007). GFGPG loop which is highly conserved in plant type III PKS was also found to be replaced by AFGPG (Fig. 5a). However, the first residue of the loop, glycine was found to be substituted with alanine (A) or serine (S) (Mallika et al., 2011). AFGPG loop was also found in curcuminoid synthase of *Oryza sativa* L.

Phylogenetic analysis showed that ClPKS11 is located together with type III PKS of *M. acuminata* in a cluster containing non-CHS enzymes including isoforms of curcumin synthase, diketide synthase, curcumin synthase of *O. sativa* etc (Fig. 5b). The amino acid sequence of ClPKS11 was highly conserved in *C. longa* irrespective of their variations in curcumin content and in other *Curcuma* spp. as in case of curcumin synthases and diketide CoA synthase (Kita et al., 2016). The secondary structure of ClPKS11 contained α-helices as the predominant component (Fig. 6). The 3D structure of ClPKS11 protein model (Fig. 7a) was confirmed using Ramachandran Plot statistics. The ClPKS11 protein model exhibited a good fit with the reference geometry with 100% of non-glycine and non-

```
Ms-CHS/1-389     1 --MVSVSEIRNAQRAEGPATTLAIGTANPTNCVEQSTYPDFYFKITNSEHKTELKEKFQRMCDKSMIKRRYMYLTEEILKENPSVCEIMAPSLDAWQDM 97
Cl-DCS/1-389     1 ---MEANGYRITHSADGPATILAIGTANPTNVVDQNAYPDFYFRVTNSEYLQELKAKFRRICEKAAIRKRHLYLTEEILRENPSLLAPMAPSFDARQAI 96
Cl-PKS11/1-391   1 MASNYVDAFPKPQRAQGPATVMAIGTANPPNLYEQSAYPDFYFRVTGADHKPELKQKFRRLCDRSMIKKRYMHLTEELLKEKPGMCSYMDTSFDERQDV 99
Cl-CURS2/1-391   1 MAMISLQAMRKAQRAQGPATILAVGTANPPNLYEQDTYPDYYFRVTNSEHKQELKNKFRLMCEKTMVKRRYLYLTPEILKERPKLCSYMEPSFDDRQDI 99
Cl-CURS1/1-389   1 --MANLHALRREQRAQGPATIMAIGTATPPNLYEQSTFPDFYFRVTNSDDKQELKKKFRRMCEKTMVKKRYLHLTEEILKERPKLCSYKEASFDDRQDI 97
Cl-CURS3/1-390   1 --MGSLQAMRRAQRAQGPATIMAVGTSNPPNLYEQTSYPDFYFRVTNSDHKHELKNKFRVICEKTKVKRRYLHLTEEILKQRPKLCSYMEPSFDDRQDI 97

Ms-CHS/1-389    98 VVVEVPRLGKEAAVKAIKEWGQPKSKITHLIVCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEVTAV 196
Cl-DCS/1-389    97 VVEAVPKLAKEAAEKAIKEWGRPKSDITHLVFCSASGIDMPGSDLQLLKLLGLPPSVNRVMLYNVGCHAGGTALRVAKDLAENNRGARVLAVCSEVTVL 195
Cl-PKS11/1-391 100 VVEEVPRLAKEAAVKAIKEWGRPLSEITHLVFCSTSGVDMPGADYRLAKLLGLSFSVNRIMLYNQACHIGAQTLRIAKDIAENNRSARVLVVACEVNTL 198
Cl-CURS2/1-391 100 VVEEVPKLAAEAAENAIKEWGGDKSAITHLVFCSISGIDMPGADYRLAQLLGLPLAVNRLMLYSQACHMGAAMLRIAKDIAENNRSARVLVVACEITVL 198
Cl-CURS1/1-389  98 VVEEIPRLAKEAAEKAIKEWGRPKSEITHLVFCSISGIDMPGADYRLATLLGLPLTVNRLMIYSQACHMGAAMLRIAKDLAENNRGARVLVVACEITVL 196
Cl-CURS3/1-390  98 VVEEIPKLAKEAAEKAIKEWGRPKSEITHLVFCSISGIDMPGADYRLATLLGLPLSVNRLMLYSQACHMGAQMLRIAKDLAENNRGARVLAVSCEITVL 196

Ms-CHS/1-389   197 TFRGPSDTHLDSLVGQALFGDGAAALIVGSDPVPEIEKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFHLLKDVPGIVSKNINKALVEAFEPLGISDY 295
Cl-DCS/1-389   196 SYRGPHPAHIESLFVQALFGDGAAALVVGSDPVDGVERPIFEIASASQVMLPESAEAVGGHLREIGLTFHLKSQLPSIIASNIEQSLTTACSPLGLSDW 294
Cl-PKS11/1-391 199 IFRGPEERDFQSLAAQVAFGDGAAALVVGADPVQGVEKPIFEIMAAFPFTVPETQMAVGGQLKQIGLTFHFAHQLPGLIANNLETCLGEALKPLGISDW 297
Cl-CURS2/1-391 199 SFRGPDERDFQALAGQAGFGDGAGAMIVGADPVLGVERPLYHIMSATQTTVPESEKAVGGHLREVGLTFHFFNQLPAIIADNVGNSLAEAFEPIGIKDW 297
Cl-CURS1/1-389 197 SFRGPNEGDFEALAGQAGFGDGAGAVVVGADPLEGIEKPIYEIAAAMQETVAESQGAVGGHLRAFGWTFYFLNQLPAIIADNLGRSLERALAPLGVREW 295
Cl-CURS3/1-390 197 SFRGPDAGDFEALACQAGFGDGAAAVVVGADPLPGVERPIYEIAAAMQETVPESERAVGGHLREIGWTFHFFNQLPKLIAENIEGSLARAFKPLGISEW 295

Ms-CHS/1-389   296 NSIFWIAHPGGPAILDQVEQKLALKPEKMKATREVLSEYGNMSSACVLFILDEMRKKSAQDGLKTTGEGLEWGVLFGFGPGLTIETVVLRSVAI-  389
Cl-DCS/1-389   295 NQLFWAVHPGGRAILDQVEARLGLEKDRLAATRHVLSEYGNMQSATVLFILDEMRNRSAAEGHATTGEGLDWGVLLGFGPGLSIETVVLHSCRLN  389
Cl-PKS11/1-391 298 NDVFWVAHPGNWGIMDAVEAKLGLEQGKLQSSRHVFSEFGNMMSATVLFVMDDVRKRAVAEGAATTGDGLKWGVLCAFGPGLSIETLVLRSVPL-  391
Cl-CURS2/1-391 298 NNIFWVAHPGNWAIMDAIETKLGLEQSKLATARHVFSEFGNMQSATVYFVMDELRKRSAAENRATTGDGLRWGVLFGFGPGISIETVVLQSVPL-  391
Cl-CURS1/1-389 296 NDVFWVAHPGNWAIIDAIEAKLQLSPDKLSTARHVFTEYGNMQSATVYFVMDELRKRSAVEGRSTTGDGLQWGVLLGFGPGLSIETVVLRSMPL-  389
Cl-CURS3/1-390 296 NDVFWVAHPGNWGIMDAIETKLGLEQGKLATARHVFSEYGNMQSATVYFVMDEVRKRSAAEGRATTGEGLEWGVLFGFGPGLTIETVVLRSVPLP  390
```

(a)

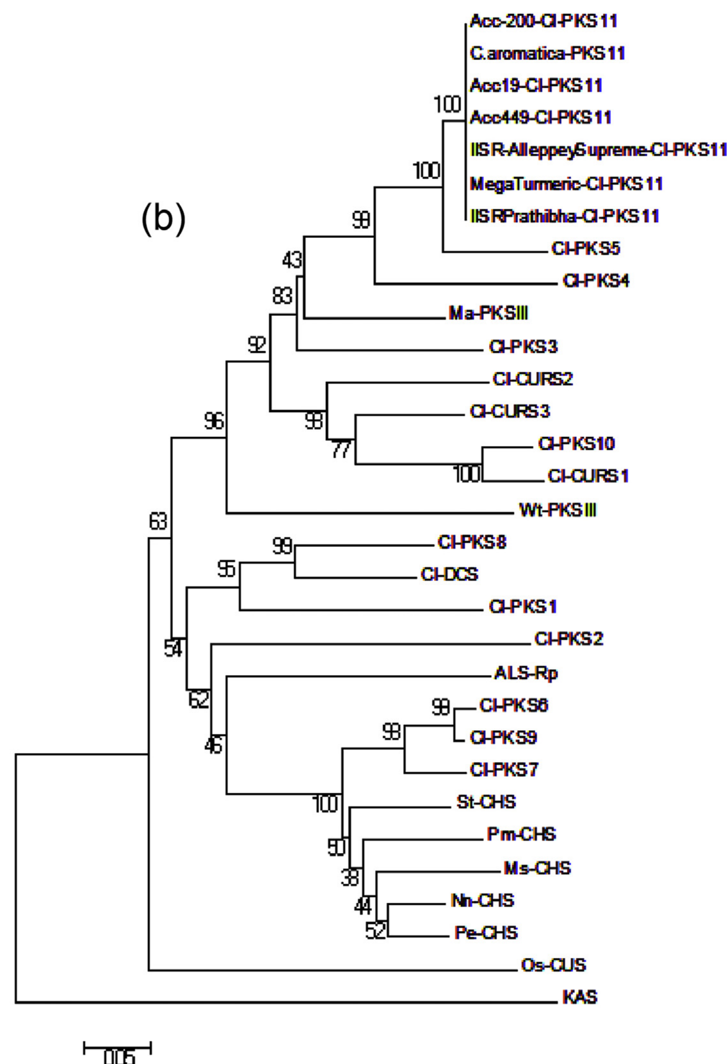


Fig. 5. Comparison of *ClPKS11* with other type III *PKS* by (a) Multiple sequence alignment and (b) Phylogenetic analysis.

proline residues suggesting that ClPKS11 protein model represents a valid stereochemical conformation (Fig. 7b). The cavity volume of ClPKS11 was predicted to be 17 Å, which is similar to that of CURS1 (18 Å).

3.6. Protein-ligand interaction

Docking results showed that all the substrates docked satisfactorily to the enzyme active site of ClPKS11 and CURS1 with

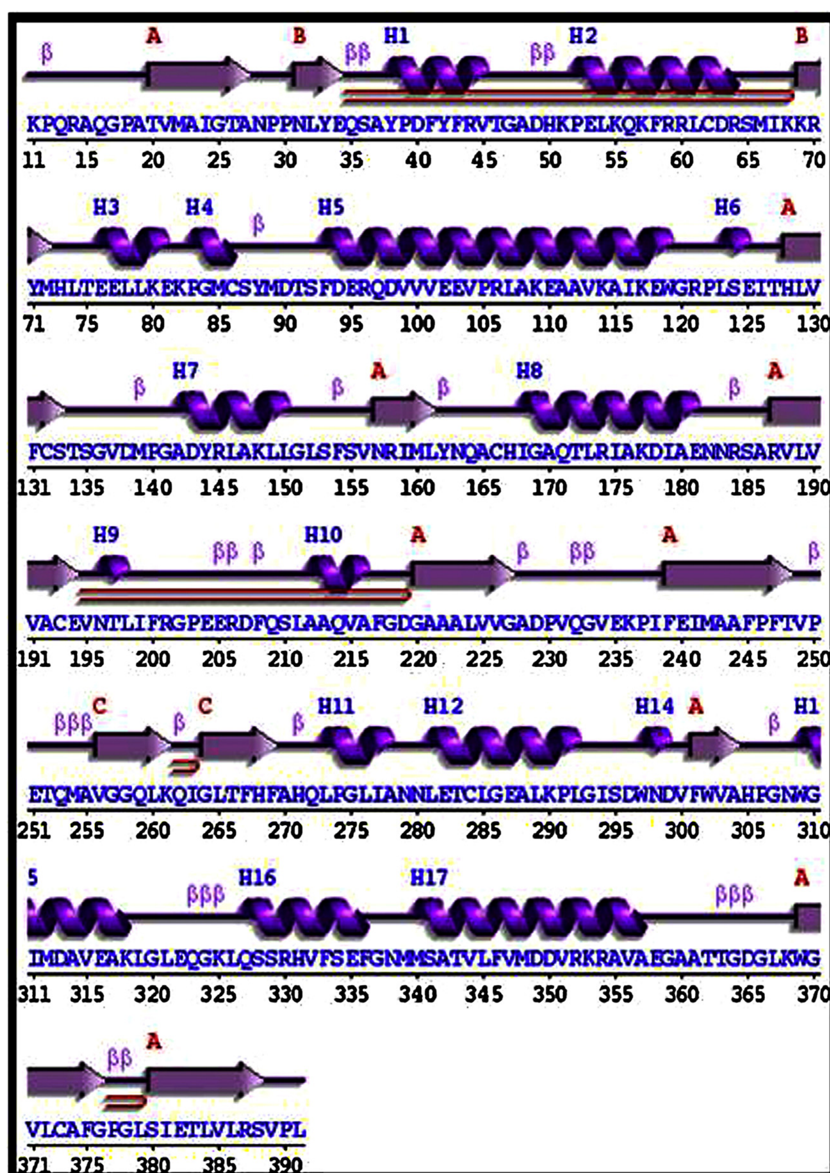


Fig. 6. Secondary structure of ClPKS11.

Table 6
Characteristics of substrates docked with ClPKS11 and CURS1 of *C. longa*.

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

good docking scores (Table 6). Molegro Virtual Docker uses an energy-based scoring function; lower energy scores represent better protein-ligand bindings compared with higher energy values (Thomsen and Christensen, 2006). Among the substrates, feruloyl CoA exhibited the lowest docking score, suggesting feruloyl CoA can be starter substrate for ClPKS11 (Fig. 8). This is agreeing with *in vitro* studies by Katsuyama et al., 2009a; where curcumin was detected even from starter substrates (*p*-coumaroyl CoA/feruloyl CoA) and diketide CoAs when incubated with purified recombinant CURS and malonyl CoA as extender. The lowest MolDock scoring function for all four substrates during the docking procedure for ClPKS11 than CURS1, suggest a higher affinity of these substrates to ClPKS11 than CURS1.

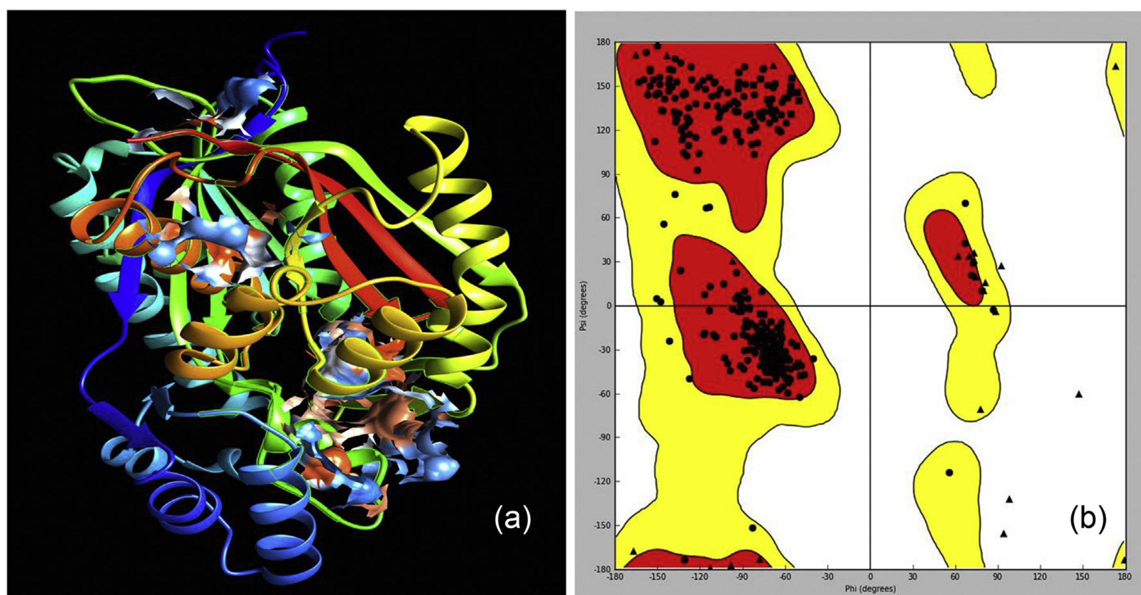


Fig. 7. Molecular modelling of *C. longa* ClPKS11 (a) 3-D structure and (b) Ramachandran plot determined by PROCHECK.

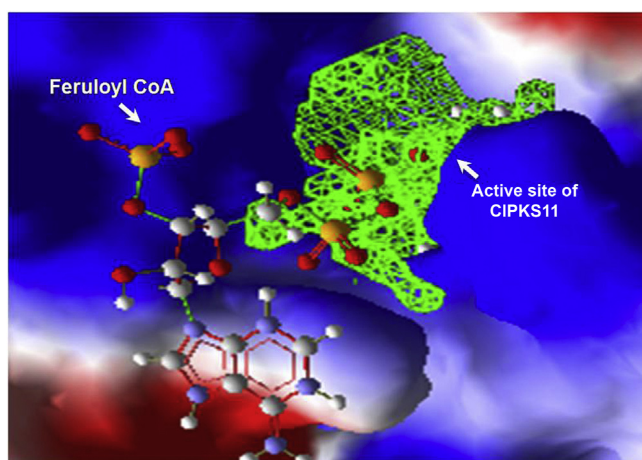


Fig. 8. Docking interaction of feruloyl coA (ball and stick model) with ClPKS11 cavities highlighted in green colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Conclusion

This is the first report establishing a direct correlation between the expression of important group of polyketide synthase genes and curcumin content in turmeric at molecular level. A highly conserved *ClPKS11* encoding type III polyketide synthase showing positive correlation with curcumin content has been isolated from turmeric and found to be novel. Its putative role in curcumin biosynthesis was confirmed by comparative transcriptome analysis and expression profiling, revealing higher fold expression corresponding to curcumin content, significant positive correlation with curcumin content, blastp analysis showing similarity to curcumin synthases, molecular docking studies with good docking score and grouping with non-CHS cluster in phylogenetic analysis. The novelty of *ClPKS11* was discussed based on blastp analysis showing maximum homology of 72% with the closest PKS and uniqueness in substrate binding and cyclization pocket residues. The characterized gene could be used to screen the germplasm for elite high curcumin varieties and will be a boon to curcumin industry.

Acknowledgements

Director, ICAR-IISR for providing the facilities.

AICRP on Spices, Tamil Nadu Agricultural University, Coimbatore for assistance in field trials.

References

Abe, I., Morita, H., 2010. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. Nat. Prod. Rep. 27, 809–838.

Abe, I., Watanabe, T., Lou, W., Noguchi, H., 2006. Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. FEBS J. 273, 208–218, http://dx.doi.org/10.1111/j.1742-4658.2005.05059.x.

Abe, I., Morita, H., Oguro, S., Noma, H., Wanibuchi, K., Kawahara, N., Goda, Y., Noguchi, H., Kohno, T., 2007. Structure-based engineering of a plant type III polyketide synthase: formation of an unnatural nonaketide naphthopyrone. J. Am. Chem. Soc. 129, 5976–5980, http://dx.doi.org/10.1021/ja070375l.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

American Spice Trade Association (ASTA), 1968. Official Analytical Methods, 2nd ed. American Spice Trade Association, New York.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., Mohanty, D., 2010. SBSPKS: Structure based sequence analysis of polyketide synthases. Nucleic Acids Res. 38, 487–496, http://dx.doi.org/10.1093/nar/gkq340.

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. Crop. Prod. 53, 358–364, http://dx.doi.org/10.1016/j.indcrop.2014.01.005.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535–542.

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

Dangelmayr, B., Stotz, G., Spribille, R., Forkmann, G., 1983. Relationship between flower development, anthocyanin accumulation and activity of enzymes involved in flavonoid biosynthesis in Matthiola incana R. Br Zeitschrift fur Naturforsch. – Sect. C J. Biosci. 38, 551–555, http://dx.doi.org/10.1515/znc-1983-7-810.

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 20, 1–9, http://dx.doi.org/10.1007/s12298-013-0218-y.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J., 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34, 116–118, http://dx.doi.org/10.1093/nar/gkl282.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinf., http://dx.doi.org/10.1002/0471250953.bi0506s15.

Felsenstein, J., 1993. Phylogeny Inference Package (PHYLIP) Version 3.5. University of Washington, Seattle.

Feng, S., Wang, Y., Yang, S., Xu, Y., Chen, X., 2010. Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. Planta 232, 245–255, http://dx.doi.org/10.1007/s00425-010-1170-5.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, Amos, 2005. Protein identification and analysis tools on the ExPASy server. Proteomics Protoc. Handb., 571–607, http://dx.doi.org/10.1385/1592598900.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95 /98/NT. Nucleic Acids Symp. Ser., 95/ (citeulike-article-id:691774).

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

Huang, C.C., Couch, G.S., Pettersen, E.F., Ferrin, T.E., 1996. Chimera: an extensive molecular modeling application constructed using standard components. Pac. Symp. Biocomput. 1, 724.

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y.-H., Yu, J.-Q., Chen, Z., 2010. Functional analysis of the Arabidopsis *PAL* gene family in plant growth, development, and response to environmental stress. Plant Physiol. 153, 1526–1538, http://dx.doi.org/10.1104/pp.110.157370.

Huang, W., Zeng, S., Xiao, G., Wei, G., Liao, S., Chen, J., Sun, W., Lv, H., Wang, Y., 2015. Elucidating the biosynthetic and regulatory mechanisms of flavonoid-derived bioactive components in *Epimedium sagittatum*. Front. Plant Sci. 6, 689, http://dx.doi.org/10.3389/fpls.2015.00689.

Jeong, S.T., Goto-Yamamoto, N., Kobayashi, S., Esaka, M., 2004. Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skins. Plant Sci. 167, 247–252, http://dx.doi.org/10.1016/j.plantsci.2004.03.021.

Jez, J.M., Austin, M.B., Ferrer, J.L., Bowman, M.E., Schröder, J., Noel, J.P., 2000. Structural control of polyketide formation in plant-specific polyketide synthases. Chem. Biol. 7, 919–930, http://dx.doi.org/10.1016/S1074-5521(00)00041-7.

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. 284, 11160–11170, http://dx.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. 583, 2799–2803, http://dx.doi.org/10.1016/j.febslet.2009.07.029.

Katsuyama, Y., Hirose, Y., Funa, N., Ohnishi, Y., Horinouchi, S., 2010. Precursor-directed biosynthesis of curcumin analogs in *Escherichia coli*. Biosci. Biotechnol. Biochem. 74, 641–645, http://dx.doi.org/10.1271/bbb.90866.

Kim, S., Han, L., Yu, B., Hahhnke, V.D., Bolton, E.E., Bryant, S.H., 2015. PubChem structure-activity relationship (SAR) clusters. J. Cheminform. 7, 1–22, http://dx.doi.org/10.1186/s13321-015-0070-x.

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. 72, 1789–1798, http://dx.doi.org/10.1271/bbb.80075.

Kita, T., Komatsu, K., Zhu, S., Iida, O., Sugimura, K., Kawahara, N., Taguchi, H., Masamura, N., Cai, S.Q., 2016. Development of intron length polymorphism markers in genes encoding diketide-CoA synthase and curcumin synthase for discriminating *Curcuma* species. Food Chem. 194, 1329–1336, http://dx.doi.org/10.1016/j.foodchem.2015.08.034.

Koo, H.J., McDowell, E.T., Ma, X., Greer, K.A., Kapteyn, J., Xie, Z., Descour, A., Kim, H., Yu, Y., Kudrna, D., Wing, R.A., Soderlund, C.A., Gang, D.R., 2013. Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. BMC Plant Biol. 13, 27, http://dx.doi.org/10.1186/1471-2229-13-27.

Kuniyoshi, H., Fukui, Y., Sakai, Y., 2006. Cloning of full-length cDNA of teleost corticotropin-releasing hormone precursor by improved inverse PCR. Biosci. Biotechnol. Biochem. 70, 1983–1986, http://dx.doi.org/10.1271/bbb.60067.

Kuttan, R., Bhanumathy, P., Nirmala, K., George, M.C., 1985. Potential anticancer activity of turmeric (*Curcuma longa*). Cancer Lett. 129 (2), 197–202.

Laskowski, R., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291, http://dx.doi.org/10.1107/S0021889892009944.

Li, L., Ban, Z., Li, X., Wu, M., Wang, A., Jiang, Y., Jiang, Y., 2012. Differential expression of anthocyanin biosynthetic genes and transcription factor *PcMYB10* in Pears (*Pyrus communis* L.). PLoS One 7, http://dx.doi.org/10.1371/journal.pone.0046070.

Li, Y., Kim, J.I., Pysh, L., Chapple, C., 2015. Four isoforms of *Arabidopsis thaliana* 4-coumarate: coA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol. 169, 00838, http://dx.doi.org/10.1104/pp.15.00838.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-[Delta][Delta] CT method. Methods 25, 402–408, http://dx.doi.org/10.1006/meth.2001.1262.

Lu, S., Zhou, Y., Li, L., Chiang, V.L., 2006. Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. Plant Cell Physiol. 47, 905–914, http://dx.doi.org/10.1093/pcp/pcj063.

Mallika, V., Sivakumar, K.C., Soniya, E.V., 2011. Evolutionary implications and physicochemical analyses of selected proteins of type iii polyketide synthase family. Evol. Bioinf. 2011, 41–53, http://dx.doi.org/10.4137/EBO.S6854.

Matsuda, S., Vert, J., Saigo, H., Ueda, N., Toh, H., Akutsu, T., 2005. A novel representation of protein sequences for prediction of subcellular location using support vector machines. Protein Sci., 2804–2813, http://dx.doi.org/10.1110/ps.051597405.

Nair, R.B., Xia, Q., Kartha, C.J., Kurylo, E., Hirji, R.N., Datla, R., Selvaraj, G., 2002. Arabidopsis CYP98A3 mediating aromatic 3- hydroxylation. Developmental regulation of the gene, and expression in yeast. Society 130, 210–220, http://dx.doi.org/10.1104/pp.008649.210.

Neema, A., 2005. Investigations on the Biosynthesis of Curcumin in Turmeric (*Curcuma Longa* L.). University of Calicut at Kerala, India (Ph.D dissertation).

Negi, P.S., Jayaprakasha, G.K., Jagan Mohan Rao, L., Sakariah, K.K., 1999. Antibacterial activity of turmeric oil: a byproduct from curcumin manufacture. J. Agric. Food Chem. 47 (10), 4297–4300.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.

Ramirez-Ahumada, M., del, C., Timmermann, B.N., Gang, D.R., 2006. Biosynthesis of curcuminoids and gingerols in turmeric (*Curcuma longa*) and ginger (*Zingiber officinale*): Identification of curcuminoid synthase and hydroxycinnamoyl-CoA thioesterases. Phytochemistry 67, 2017–2029, http://dx.doi.org/10.1016/j.phytochem.2006.06.028.

Resmi, M.S., Soniya, E.V., 2012. Molecular cloning and differential expressions of two cDNA encoding Type III polyketide synthase in different tissues of *Curcuma longa* L. Gene 491, 278–283, http://dx.doi.org/10.1016/j.gene.2011.09.025.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the european molecular biology open software suite. Trends Genet. 16, 276–277, http://dx.doi.org/10.1016/j.cocis.2008.07.002.

Rombel, I.T., Sykes, K.F., Rayner, S., Johnston, S.A., 2002. ORF-FINDER: A vector for high-throughput gene identification. Gene 282, 33–41, http://dx.doi.org/10.1016/S0378-1119(01)00819-8.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574, http://dx.doi.org/10.1093/bioinformatics/btg180.

Roughley, P.J., Whiting, D.A., 1973. Experiments in the biosynthesis of curcumin. J. Chem. Soc. Perkin Trans. 1, 2379–2388.

Sadowski, J., Gasteiger, J., Schwab, C.H., 2003. 3D structure generation and conformational searching. In: Computational Medicinal Chemistry and Drug Discovery. Dekker Inc, New York, NY, pp. 151–212.

Sajitha, P.K., Prasath, D., Sasikumar, B., 2014. Phenological variation in two species of *Curcuma*. 42, 252–255.

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. 10, 227–240, http://dx.doi.org/10.1007/s11816-016-0399-2.

Sasikumar, B., 2005. Genetic resources of *Curcuma*: diversity, characterization and utilization. Plant Genet Res. 3, 230–251.

Satoskar, R.R., Shah, S.J., Shenoy, S.G., 1986. Evaluation of anti-inflammatory property of curcumin (diferuloyl methane) in patients with postoperative inflammation International Journal of Clinical Pharmacology. Ther. Toxicol. 24 (12), 651–654.

Schröder, J., 1997. A family of plant-specific polyketide synthases: facts and predictions. Trends Plant Sci. 2, 373–378, http://dx.doi.org/10.1016/S1360-1385(97)01104-7.

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. Rep., 1825–1836, http://dx.doi.org/10.1007/s11105-015-0878-6.

Shi, S.G., Yang, M., Zhang, M., Wang, P., Kang, Y.-X., Liu, J.-J., 2014. Genome-wide transcriptome analysis of genes involved in flavonoid biosynthesis between red and white strains of *Magnolia sprengeri* pamp. BMC Genomics 15, 706, http://dx.doi.org/10.1186/1471-2164-15-706.

Sidhu, G.S., Singh, A.K., Thaloor, D., Banaudha, K.K., Patnaik, G.K., Srimal, R.C., Maheshwari, R.K., 1998. Enhancement of wound healing by curcumin in animals. Wound Repair Regen. 6 (2), 167–177.

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. 43, 165–171, http://dx.doi.org/10.1016/j.indcrop.2012.07.006.

Steyn, W.J., Wand, S.J.E., Holcroft, D.M., Jacobs, G., 2005. Red colour development and loss in pears. Acta Hortic. 671, 79–85.

Sun, X., Gong, S.Y., Nie, X.Y., Li, Y., Li, W., Huang, G.Q., Li, X.-B., 2015. A R2R3-MYB transcription factor that is specifically expressed in cotton (*Gossypium hirsutum*) fibers affects secondary cell wall biosynthesis and deposition in transgenic Arabidopsis. Physiol. Plant 154, 420–432, http://dx.doi.org/10.1111/ppl.12317.

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. 22, 4673–4680, http://dx.doi.org/10.1093/nar/22.22.4673.

Thomsen, R., Christensen, M.H., 2006. MolDock: a new technique for high accuracy docking. J. Med. Chem. 49, 3315–3321.

Toda, S., Miyase, T., Arichi, H., Tanizawa, H., Takino, Y., 1985. Natural antioxidants. III. Antioxidative components isolated from rhizome of *Curcuma longa* L. Chem. Pharm. Bull. 33 (4), 1725–1728.

Ubi, B.E., Honda, C., Bessho, H., Kondo, S., Wada, M., Kobayashi, S., Moriguchi, T., 2006. Expression analysis of anthocyanin biosynthetic genes in apple skin: effect of UV-B and temperature. Plant Sci. 170, 571–578, http://dx.doi.org/10.1016/j.plantsci.2005.10.009.

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, 746, http://dx.doi.org/10.1186/1471-2164-15-746.

Wei, Y.Z., Hu, F.C., Hu, G.B., Li, X.J., Huang, X.M., Wang, H.C., 2011. Differential expression of anthocyanin biosynthetic genes in relation to anthocyanin accumulation in the pericarp of litchi chinensis sonn. PLoS One 6, http://dx.doi.org/10.1371/journal.pone.0019455.

Xie, Z., Ma, X., Gang, D.R., 2009. Modules of co-regulated metabolites in turmeric (*Curcuma longa*) rhizome suggest the existence of biosynthetic modules in plant specialized metabolism. J. Exp. Bot. 60, 87–97, http://dx.doi.org/10.1093/jxb/ern263.

Xie, F., Xiao, P., Chen, D., Xu, L., Zhang, B., 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol. Biol. 80, 75–84, http://dx.doi.org/10.1007/s11103-012-9885-2.

Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion. The University of Texas at Austin (Ph.D. dissertation).